Testing linearity against threshold effects: uniform inference in quantile regression

Antonio F. Galvao, Jr.[*] Kengo Kato[†] Gabriel Montes-Rojas[‡] Jose Olmo[§]

February 5, 2013

Abstract

This paper develops a uniform test of linearity against thresholds effects in the quantile regression framework. The test is based on the supremum of the Wald process over the space of quantile and threshold parameters. We establish the limiting null distribution of the test statistic for stationary weakly dependent processes, and propose a simulation method to approximate the critical values. The proposed simulation method makes the test easy to implement. Monte Carlo experiments show that the proposed test has good size and reasonable power against nonlinear threshold models.

Key words: Linearity test, quantile regression, threshold model.

1 Introduction

This paper develops a uniform test of linearity against threshold nonlinearity of the conditional quantile function for stationary time series processes. The null hypothesis assumes that the conditional quantile function is linear in the conditioning variables uniformly over a given range of quantiles, while the alternative hypothesis assumes

[*]Department of Economics, University of Iowa, W210 Pappajohn Business Building, 21 E. Market Street, Iowa City, IA 52242; and University of Wisconsin-Milwaukee, Bolton Hall 852, 3210 N. Maryland Ave., Milwaukee, WI 53201 E-mail: `antonio-galvao@uiowa.edu`

[†]Corresponding Author: Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan. E:mail: `kkato@hiroshima-u.ac.jp`

[‡]Department of Economics, City University London, D306 Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK. Email: `Gabriel.Montes-Rojas.1@city.ac.uk`

[§]Centro Universitario de la Defensa and ARAID, Academia General Militar, Ctra. Huesca s/n, 50003, Zaragoza, Spain. Email: `jolmo@unizar.es`

that the conditional quantile function follows a threshold model at some quantile, i.e., is piecewise linear in the conditioning variables at some quantile. Under this formulation, we develop a test based on the supremum of the Wald process over the space of quantile and threshold parameters. We establish the limiting null distribution of the proposed test for stationary weakly dependent processes, and show that it is consistent against fixed alternatives and has nontrivial power against the Pitman family of local alternatives. Unfortunately, the limiting null distribution is not pivotal because it depends on unknowns. In order to compute critical values of the test, we propose a simple simulation method that exploits the specific property of the quantile regression estimator and establish its validity. We carry out Monte Carlo experiments to study the finite sample properties of the proposed test in terms of empirical size and power. The results show evidence that the proposed test presents empirical size very close to the nominal size, and reasonable power performance. The simulation exercise also confirms in finite samples the nontrivial power of our test statistic against the Pitman family of local alternatives.

There is a large number of studies on threshold models, in particular in the time series literature (Tong and Lim, 1980; Tong, 1983; Tsay, 1989; Chan, 1990; Hansen, 1996, 1997, 2000), and on threshold quantile regression as well (Caner, 2002; Cai and Stander, 2008; Cai, 2010). This paper contributes to the literature by proposing a convenient testing procedure on threshold effects, and by studying its theoretical and practical performance in the quantile regression framework. We shall comment that threshold models may act as a general alternative to the null of linearity. In fact, Fan and Yao (2005, p.134) stated: "Although the test is designed for a specified alternative, it may be applied to test a departure to a general smooth nonlinear function since a piecewise linear function will provide a better approximation than that from a (global) linear function."

Quantile regression is applied in many fields because it allows for statistical inference on the entire conditional distribution. Many researchers have investigated the problem of testing the hypothesis of linearity of the conditional quantile function (Zheng, 1998; Horowitz and Spokoiny, 2002; He and Zhu, 2003; Whang, 2005; Otsu, 2008). These tests are pointwise, i.e., they test linearity of the conditional quantile function at a given quantile, say median, and their formulations of the testing problems are different from ours. In an independent study, Lee et al. (2011) considered a test of linearity against a threshold nonlinear alternative for a fixed quantile. This test is also pointwise, and they do not allow for dynamics in the model. Therefore, the uniform test proposed in this paper differs from those available in the literature. We shall stress that uniform inference has an important role in the quantile regression literature. For

a related literature on uniform inference in quantile regression, we refer to Gutenbrunner and Jurečková (1992), Koenker and Machado (1999), Koenker and Xiao (2002), Chernozhukov and Fernández-Val (2005), and Angrist et al. (2006). The last paper states advantages of uniform inference over pointwise one in some detail. It is important to note that in our case it is difficult to control the size of the overall procedure if one applies pointwise tests for a number of quantile indices. More recently, Escanciano and Velasco (2010) proposed general specification tests of parametric dynamic quantile regression models in a different way than ours.

In the quantile regression framework, Qu (2008) and Su and Xiao (2008) developed uniform tests for structural changes. However, there is an important difference between threshold and structural break models. In the former, the nonlinearity is defined by the observable history of the time series, and in the latter, the conditional distribution changes at an exogenous date. Moreover, as Carrasco (2002) argued, tests for structural change have no power if the data are generated by Markov-switching or threshold models.

The paper is organized as follows. In Section 2, we formally define the test statistic. In Section 3, we establish its limiting nulll distribution and show its consistency. We also briefly comment on the limiting distribution of the test statistic under local alternatives. In Section 4, we develop a simulation method to approximate the critical values of the proposed test. In Section 5, we present Monte Carlo experiments of the test's finite sample performance. In Section 6, we give a brief summary of the paper. Proofs are gathered in the Appendix.

2   Formulation and test statistic

Let $(y_t, q_t, \boldsymbol{x}_t')'$ be a triple of a scalar dependent variable y_t, a scalar threshold variable q_t and a vector of d explanatory variables $\boldsymbol{x}_t$ that may contain lags of y_t. A typical example of q_t is an element of $\boldsymbol{x}_t$. Define $\boldsymbol{z}_t := (q_t, \boldsymbol{x}_t')'$. Throughout the paper, we assume that the process $\{(y_t, \boldsymbol{z}_t')', t \in \mathbb{Z}\}$ is strictly stationary. Let $\mathcal{A}_{t-1}$ denote the σ-field generated by $\{\boldsymbol{z}_t, y_{t-1}, \boldsymbol{z}_{t-1}, y_{t-2}, \dots\}$ and let $Q_{y_t}(\tau|\mathcal{A}_{t-1})$ denote the conditional τ-quantile of y_t given $\mathcal{A}_{t-1}$, where $\tau \in (0, 1)$.

We consider testing the null hypothesis

$$H_0 : Q_{y_t}(\tau|\mathcal{A}_{t-1}) = \boldsymbol{x}_t'\boldsymbol{\theta}_1(\tau), \text{ for all } \tau \in \mathcal{T}$$

against the alternative

$$H_1 : Q_{y_t}(\tau_0|\mathcal{A}_{t-1}) = I(q_t > \gamma_0)\boldsymbol{x}_t'\boldsymbol{\theta}_1(\tau_0) + I(q_t \leq \gamma_0)\boldsymbol{x}_t'\boldsymbol{\theta}_2(\tau_0), \quad \text{for some } \tau_0 \in \mathcal{T},$$

where $I(\cdot)$ is the indicator function, $\mathcal{T} := [\tau_L, \tau_U]$ is a bounded closed interval in $(0, 1)$ and γ_0 is the threshold parameter. Let $\Gamma := [\gamma_L, \gamma_U]$ be the parameter space of γ_0. The null hypothesis assumes that the conditional quantile function is linear in $\boldsymbol{x}_t$ uniformly over a given range of quantiles, while the alternative hypothesis assumes that the conditional quantile function follows a threshold model at some quantile. To differentiate the alternative from the null hypothesis, we assume that $\boldsymbol{\theta}_1(\tau_0) \neq \boldsymbol{\theta}_2(\tau_0)$.

It will be convenient to write the hypotheses in a different form. Let $\boldsymbol{\beta}_{(1)}(\tau_0) = \boldsymbol{\theta}_1(\tau_0)$ and $\boldsymbol{\beta}_{(2)}(\tau_0) = \boldsymbol{\theta}_2(\tau_0) - \boldsymbol{\theta}_1(\tau_0)$. Then the alternative hypothesis is expressed as

$$H_1 : Q_{y_t}(\tau_0|\mathcal{A}_{t-1}) = \boldsymbol{z}_t(\gamma_0)'\boldsymbol{\beta}(\tau_0) \text{ with } \boldsymbol{\beta}_{(2)}(\tau_0) \neq \boldsymbol{0}, \text{ for some } \tau_0 \in \mathcal{T},$$

where $\boldsymbol{z}_t(\gamma) = (\boldsymbol{x}_t', I(q_t \leq \gamma)\boldsymbol{x}_t')'$ and $\boldsymbol{\beta}(\tau_0) = (\boldsymbol{\beta}_{(1)}(\tau_0)', \boldsymbol{\beta}_{(2)}(\tau_0)')'$. Working with this notation, we may write the null hypothesis as

$$H_0 : Q_{y_t}(\tau|\mathcal{A}_{t-1}) = \boldsymbol{z}_t(\gamma)'\boldsymbol{\beta}(\tau) \text{ with } \boldsymbol{\beta}_{(2)}(\tau) = \boldsymbol{0}, \text{ for all } \tau \in \mathcal{T}, \tag{1}$$

regardless of the value of $\gamma \in \Gamma$. These alternative expressions of the null and the alternative hypotheses lead to the following testing procedure: suppose that the sample $\{(y_t, \boldsymbol{z}_t')'\}_{t=1}^n$ is given. Given $(\tau, \gamma) \in \mathcal{T} \times \Gamma$, let $\hat{\boldsymbol{\beta}}(\tau, \gamma)$ be the estimator defined by

$$\hat{\boldsymbol{\beta}}(\tau, \gamma) := \arg \min_{\boldsymbol{b} \in \mathbb{R}^{2d}} \frac{1}{n} \sum_{t=1}^n \rho_\tau(y_t - \boldsymbol{z}_t(\gamma)'\boldsymbol{b}),$$

where $\rho_\tau(u) := u\{\tau - I(u \leq 0)\}$ is the check function (Koenker and Bassett, 1978). This $\hat{\boldsymbol{\beta}}(\tau, \gamma)$ is the quantile regression estimator when we treat $\boldsymbol{z}_t(\gamma)$ as "explanatory variables". When H_0 is true, under suitable regularity conditions, $\hat{\boldsymbol{\beta}}_2(\tau, \gamma)$ converges in probability to $\boldsymbol{0}$ for each $(\tau, \gamma) \in \mathcal{T} \times \Gamma$. On the other hand, when H_1 is true, $\hat{\boldsymbol{\beta}}_2(\tau_0, \gamma_0)$ converges in probability to $\boldsymbol{\beta}_{(2)}(\tau_0) \neq 0$. However, we know *a priori* neither the quantile τ_0 where the linearity breaks down nor the true value of the threshold parameter γ_0 at that quantile. Therefore, it is reasonable to reject H_0 if the magnitude of $\hat{\boldsymbol{\beta}}_2(\tau, \gamma)$ is suitably large for some $(\tau, \gamma) \in \mathcal{T} \times \Gamma$. A natural choice is to test H_0 against H_1 by the supremum of the Wald process

$$SW_n := \sup_{(\tau, \gamma) \in \mathcal{T} \times \Gamma} n\hat{\boldsymbol{\beta}}_{(2)}(\tau, \gamma)'\{V_{22}(\tau, \gamma)\}^{-1}\hat{\boldsymbol{\beta}}_{(2)}(\tau, \gamma), \tag{2}$$

4

where $V_{22}(\tau, \gamma)$ is the asymptotic covariance matrix of $\sqrt{n}\hat{\boldsymbol{\beta}}_2(\tau, \gamma)$ under H_0. In practice, $V_{22}(\tau, \gamma)$ is replaced by a suitable consistent estimate.

The problem of our test is that the threshold parameter is not identified under the null hypothesis. Such a problem is called the Davies problem (see Davies, 1977, 1987). For works that address the Davies problem in a general but different context, we refer to Andrews and Ploberger (1994) and Hansen (1996) among many others. The difference from the standard situation in the Davies problem is that we now take the supremum over two parameters (τ, γ) in the definition of SW_n. We remark that one could use more general functionals than taking supremum. The limiting null distribution of such functionals can be obtained by the continuous mapping theorem and the weak convergence result established in Theorem 1 below. However, a detailed treatment of their properties is beyond the scope of this paper. In what follows, we restrict our attention to the supremum functional.

3 Large sample theory

3.1 Limiting null distribution and consistency

In this subsection, we derive the limiting null distribution of the proposed test statistic SW_n and show its consistency against fixed alternatives. To this end, we derive the limiting null distribution of the two-parameter process $\hat{\boldsymbol{\beta}}(\tau, \gamma)$ on $\mathcal{T} \times \Gamma$. Here we make the following regularity conditions.

(C1) The process $\{(y_t, \boldsymbol{z}_t')', t \in \mathbb{Z}\}$ is strict stationary and β-mixing with β-mixing coefficients satisfying $\beta(j) = O(j^{-l})$ with $l > p/(p-2)$, where $p > 2$ is given in condition (C2) below.

(C2) $\mathrm{E}[\|\boldsymbol{x}_t\|^p] < \infty$ for some $p > 2$.

(C3) Let $F(\cdot|\boldsymbol{z})$ denote the conditional distribution function of y_t given $\boldsymbol{z}_t = \boldsymbol{z}$. Assume that $F(\cdot|\boldsymbol{z})$ has a Lebesgue density $f(\cdot|\boldsymbol{z})$ such that

    (i) $|f(y|\boldsymbol{z})| \leq C_f$ on the support of $(y_t, \boldsymbol{z}_t')'$ for some constant $C_f > 0$.

    (ii) $|f(y_1|\boldsymbol{z}) - f(y_2|\boldsymbol{z})| \to 0$ as $|y_1 - y_2| \to 0$ for each fixed $\boldsymbol{z}$.

(C4) The threshold variable q_t has a continuous distribution.

(C5) There exist an open set $\mathcal{T}^* \subset (0, 1)$ with $\mathcal{T}^* \supset \mathcal{T}$ such that for each $\tau \in \mathcal{T}^*$, there exists a unique vector $\boldsymbol{\beta}_{(1)}^*(\tau) \in \mathbb{R}^p$ that solves the equation $\mathrm{E}[(\tau - I\{y_t \leq \boldsymbol{x}_t'\boldsymbol{\beta}_{(1)}^*(\tau)\})\boldsymbol{x}_t] = \boldsymbol{0}$.

(C6) Define the matrices

$$\Omega_0(\gamma_1, \gamma_2) := \mathrm{E}[\boldsymbol{z}_t(\gamma_1)\boldsymbol{z}_t(\gamma_2)'], \Omega_1(\tau, \gamma) := \mathrm{E}[f(\boldsymbol{x}_t'\boldsymbol{\beta}_{(1)}^*(\tau)|\boldsymbol{z}_t)\boldsymbol{z}_t(\gamma)\boldsymbol{z}_t(\gamma)'].$$

Assume that $\Omega_0(\gamma, \gamma)$ is positive definite for each $\gamma \in \Gamma$, and $\Omega_1(\tau, \gamma)$ is positive definite for each $(\tau, \gamma) \in \mathcal{T} \times \Gamma$.

Condition (C1) allows for time series data. We require the process to be β-mixing. This is because our proof uses the uniform central limit theorem for β-mixing processes displayed in Arcones and Yu (1995). For some basic properties of mixing processes, we refer to Section 2.6 of Fan and Yao (2005) and references therein. Condition (C2) is a moment condition. Condition (C3) is standard in the quantile regression literature (see Angrist et al., 2006). Condition (C4) is standard in the threshold regression literature (see Hansen, 1996, 2000). Condition (C5) needs an explanation. When H_0 is true, $\boldsymbol{\beta}_{(1)}(\tau) = \boldsymbol{\beta}_{(1)}^*(\tau)$ in (1) and $Q_{y_t}(\tau|\mathcal{A}_{t-1}) = \boldsymbol{x}_t'\boldsymbol{\beta}_{(1)}^*(\tau)$ for all $\tau \in \mathcal{T}$. When H_0 is not true, $\boldsymbol{\beta}_{(1)}^*(\tau)$ is interpreted as the coefficient vector of the best linear predictor of the conditional quantile function against a certain weighted mean-squared loss function (Angrist et al., 2006, Theorem 1). The reason to assume condition (C5) is that the limiting null distribution of SW_n depends on the probability limit of $\hat{\boldsymbol{\beta}}_1(\tau, \gamma)$ under the null hypothesis. To guarantee that this distribution is well defined under the alternative hypothesis (which is relevant when we argue about consistency of the test against fixed alternatives), we need condition (C5). Under these conditions, the map $\tau \in \mathcal{T}^* \mapsto \boldsymbol{\beta}_{(1)}^*(\tau)$ is continuously differentiable by the implicit function theorem (see Angrist et al., 2006, p.560). Condition (C6) guarantees that the matrices $\Omega_0(\gamma, \gamma)$ and $\Omega_1(\tau, \gamma)$ do not degenerate for each fixed $\gamma \in \Gamma$ and $(\tau, \gamma) \in \mathcal{T} \times \Gamma$, respectively. In fact, they do not degenerate uniformly over those sets under the present conditions:

Lemma 1. *Under conditions (C2)-(C5), the map $(\gamma_1, \gamma_2) \mapsto \Omega_0(\gamma_1, \gamma_2)$ is continuous on $\Gamma \times \Gamma$, and the map $(\tau, \gamma) \mapsto \Omega_1(\tau, \gamma)$ is continuous on $\mathcal{T} \times \Gamma$.*

Proof. Follows from standard calculations. □

Because of the computational property of the quantile regression estimate (Koenker and Bassett, 1978, Theorem 3.1), we can select $\hat{\boldsymbol{\beta}}(\tau, \gamma)$ in such a way that the path $(\tau, \gamma) \mapsto \hat{\boldsymbol{\beta}}(\tau, \gamma)$ is bounded. Moreover, the path $\tau \mapsto \boldsymbol{\beta}_{(1)}^*(\tau)$ is continuous as mentioned before. Therefore, we may assume that the path $(\tau, \gamma) \mapsto \sqrt{n}\{\hat{\boldsymbol{\beta}}(\tau, \gamma) - (\boldsymbol{\beta}_{(1)}^*(\tau)', \boldsymbol{0}')'\}$ is bounded over $(\tau, \gamma) \in \mathcal{T} \times \Gamma$. Let $\ell^\infty(\mathcal{T} \times \Gamma)$ denote the space of all bounded functions on $\mathcal{T} \times \Gamma$ equipped with the uniform topology, and $(\ell^\infty(\mathcal{T} \times \Gamma))^{2d}$ denote the $(2d)$-product space of $\ell^\infty(\mathcal{T} \times \Gamma)$ equipped with the product topology. We use the notation "$\Rightarrow$" for weak convergence. We refer to Chapter 1 of van der Vaart

and Wellner (1996) for weak convergence in general non-separable metric spaces. For $a, b \in \mathbb{R}$, we write $a \wedge b = \min\{a, b\}$. Here let $\boldsymbol{\beta}^*(\tau) := (\boldsymbol{\beta}_{(1)}^*(\tau)', \mathbf{0}')' \in \mathbb{R}^{2d}$.

Theorem 1. *Assume conditions (C1)-(C6). Then under H_0 (that is, $Q_{y_t}(\tau | \mathcal{A}_{t-1}) = \boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau)$ for all $\tau \in \mathcal{T}$), $\sqrt{n}\{\hat{\boldsymbol{\beta}}(\tau, \gamma) - \boldsymbol{\beta}^*(\tau, \gamma)\}$ admits the Bahadur representation*

$$\sqrt{n}\{\hat{\boldsymbol{\beta}}(\tau, \gamma) - \boldsymbol{\beta}^*(\tau)\} = \Omega_1(\tau, \gamma)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\tau - I\{y_t \leq \boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau)\} \right] \boldsymbol{z}_t(\gamma) + r_n(\tau, \gamma), \quad (3)$$

where $\sup_{(\tau, \gamma) \in \mathcal{T} \times \Gamma} \|r_n(\tau, \gamma)\| = o_p(1)$. Therefore, under H_0,

$$\sqrt{n}\{\hat{\boldsymbol{\beta}}(\tau, \gamma) - \boldsymbol{\beta}^*(\tau)\} \Rightarrow \Omega_1(\tau, \gamma)^{-1} W(\tau, \gamma) \text{ in } (\ell^\infty(\mathcal{T} \times \Gamma))^{2d}, \quad (4)$$

where $W(\tau, \gamma)$ is a zero-mean, continuous Gaussian process on $\mathcal{T} \times \Gamma$ with covariance kernel

$$\mathrm{E}[W(\tau_1, \gamma_1) W(\tau_2, \gamma_2)'] = (\tau_1 \wedge \tau_2 - \tau_1 \tau_2) \Omega_0(\gamma_1, \gamma_2). \quad (5)$$

A proof of Theorem 1 is given in Appendix A.

Remark 1. Under the present specification, $\{\tau - I\{y_t \leq \boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau)\}, t \in \mathbb{Z}\}$ is a martingale difference sequence (m.d.s.) with respect to $\{\mathcal{A}_{t-1}, t \in \mathbb{Z}\}$ under H_0, so that the limiting null distribution of $\hat{\boldsymbol{\beta}}(\tau, \gamma)$ is the same as the one as if the observations were independent. In particular, the asymptotic covariance matrix is not an infinite sum. The m.d.s. assumption can be violated, for example, when error processes are serially dependent or of GARCH type. However, the m.d.s. assumption is widely used in time series analysis of quantile regression models (see, e.g., Qu, 2008; Escanciano and Velasco, 2010; Komunjer and Vuong, 2010a,b).[1] This is partly because estimation of long-run covariance matrices for quantile regression estimators is not well developed (note that Andrews (1991) is not applicable since the moment functions in the quantile regression case are discontinuous). For the sake of simplicity, we maintain the m.d.s. assumption under the null (note that the m.d.s. assumption is assumed only under the null, and under alternatives this assumption can be violated).

Given Theorem 1, it is now immediate to derive the limiting null distribution of the test statistic SW_n defined by (2). Let $R := [O \; I_d]$, the $d \times 2d$ matrix.

Corollary 1. *Assume conditions (C1)-(C6). Then under H_0,*

$$SW_n \Rightarrow \sup_{(\tau, \gamma) \in \mathcal{T} \times \Gamma} S(\tau, \gamma)' \{V_{22}(\tau, \gamma)\}^{-1} S(\tau, \gamma), \quad (6)$$

---

[1]An important exception is Su and White (2012). This reference was noted by a referee.

where $S(\tau, \gamma) = R\Omega_1(\tau, \gamma)^{-1}W(\tau, \gamma)$ and $V_{22}(\tau, \gamma)$ is the asymptotic covariance matrix of $\sqrt{n}\hat{\boldsymbol{\beta}}_2(\tau, \gamma)$ under H_0:

$$V_{22}(\tau, \gamma) = \mathrm{E}[S(\tau, \gamma)S(\tau, \gamma)'] = \tau(1 - \tau)R\Omega_1(\tau, \gamma)^{-1}\Omega_0(\gamma, \gamma)\Omega_1(\tau, \gamma)^{-1}R'.$$

Proof. Follows from the continuous mapping theorem and the weak convergence result established in Theorem 1. $\qquad\square$

Recall that, thanks to condition (C5), the distribution given by (6) is in general well defined. The limiting null distribution of SW_n is given by the supremum of a two-parameter chi-square process and is not pivotal. In fact, it depends on the unknown parameter $\boldsymbol{\beta}^*_{(1)}(\tau)$ and the distribution of $\boldsymbol{z}_t$, so critical values cannot be universally tabulated except for some special cases. We will discuss the practical implementation in Section 4.

It is standard to show that under the same conditions of Theorem 1, under the alternative hypothesis H_1, $\hat{\boldsymbol{\beta}}_{(2)}(\tau_0, \gamma_0) \xrightarrow{p} \boldsymbol{\beta}_{(2)}(\tau_0) \neq \boldsymbol{0}$, which implies that $SW_n \xrightarrow{p} \infty$ under H_1. So the proposed test is consistent against fixed alternatives.

3.2 Limiting distribution under local alternatives

We briefly comment on the limiting distribution of the test SW_n under local alternatives, which requires a slightly different formulation and a set of different regularity conditions. Consider the following local alternatives:

$$H_{1n} : Q_{y_{n,t}}(\tau|\mathcal{A}^n_{t-1}) = \boldsymbol{x}'_{n,t}\boldsymbol{\beta}^*_{(1)}(\tau) + I(q_{n,t} \leq \gamma_0)\boldsymbol{x}'_{n,t}\boldsymbol{\beta}_{(2),n}(\tau),$$
$$\text{with } \boldsymbol{\beta}_{(2),n}(\tau) = n^{-1/2}\boldsymbol{c}(\tau), \text{ for all } \tau \in \mathcal{T},$$

where the observations $(y_{n,t}, q_{n,t}, \boldsymbol{x}_{n,t})$ are now indexed by n (with $\mathcal{A}^n_{t-1}$ being the σ-field generated by $(q_{n,t}, \boldsymbol{x}_{n,t}), (y_{n,s}, q_{n,s}, \boldsymbol{x}_{n,s}), s \leq t - 1$), γ_0 is some point in Γ, and the maps $\tau \mapsto \boldsymbol{\beta}^*_{(1)}(\tau)$ and $\tau \mapsto \boldsymbol{c}(\tau)$ are d-vectors of continuous and bounded functions on $\mathcal{T}$, respectively. Then subject to some tecnhical conditions, which guarantee versions of (C1)-(C4) and (C6) uniformly in n, and uniform convergence of some matrices such as $\mathrm{E}[\boldsymbol{z}_{n,t}(\gamma_1)\boldsymbol{z}_{n,t}(\gamma_2)]$ (where $\boldsymbol{z}_{n,t}(\gamma) = (\boldsymbol{x}'_{n,t}, I(q_{n,t} \leq \gamma)\boldsymbol{x}'_{n,t})'$), it will be shown that under H_{1n}, SW_n (with a suitable change in $V_{22}(\tau, \gamma)$) converges in distribution to the supremum of a non-central chi-square process, as in Hansen (1996) in the least squares case. More formally, we have the following proposition. Let $\boldsymbol{z}_{n,t} = (q_{n,t}, \boldsymbol{x}'_{n,t})'$ and denote by $f_n(y_{n,t}|\boldsymbol{z}_{n,t})$ the conditional density of $y_{n,t}$ given $\boldsymbol{z}_{n,t}$ (which we assume to

exist). Moreover, let here

$$\Omega_0(\gamma_1, \gamma_2) = \lim_{n \to \infty} \mathrm{E}[\boldsymbol{z}_{n,t}(\gamma_1)\boldsymbol{z}_{n,t}(\gamma_2)], \tag{7}$$

$$\Omega_2(\tau, \gamma_1, \gamma_2) = \lim_{n \to \infty} \mathrm{E}[f_n(\boldsymbol{x}'_{n,t}\boldsymbol{\beta}^*_{(1)}(\tau)|\boldsymbol{z}_{n,t})\boldsymbol{z}_{n,t}(\gamma_1)\boldsymbol{z}_{n,t}(\gamma_2)], \ \Omega_1(\tau, \gamma) = \Omega_2(\tau, \gamma, \gamma), \tag{8}$$

where we assume the limits exist.

Proposition 1. *Assume conditions (D1)-(D5) in Appendix B. Then under H_{1n},*

$$\sqrt{n}\hat{\boldsymbol{\beta}}_{(2)}(\tau, \gamma) \Rightarrow S(\tau, \gamma) + \tilde{\boldsymbol{c}}(\tau, \gamma, \gamma_0), \ in \ (\ell^\infty(\mathcal{T} \times \Gamma))^d,$$

where $S(\tau, \gamma)$ is as in Corollary 1 (with $\Omega_0(\gamma_1, \gamma_2)$ and $\Omega_1(\tau, \gamma)$ repaced by (7) and (8)), and $\tilde{\boldsymbol{c}}(\tau, \gamma, \gamma_0) = R\Omega_1(\tau, \gamma)^{-1}\Omega_2(\tau, \gamma, \gamma_0)R'\boldsymbol{c}(\tau)$.

A proof of Proposition 1 is given in Appendix B.

4 Implementation

We now go back to the setting in Section 3.1. To implement the proposed test we have to estimate the matrices $\Omega_0(\gamma, \gamma)$ and $\Omega_1(\tau, \gamma)$. It is natural to use $\hat{\Omega}_0(\gamma_1, \gamma_2) := n^{-1}\sum_{t=1}^n \boldsymbol{z}_t(\gamma_1)\boldsymbol{z}_t(\gamma_2)'$ as an estimator of $\Omega_0(\gamma_1, \gamma_2)$. In fact, $\hat{\Omega}_0(\gamma_1, \gamma_2)$ is shown to be uniformly consistent by the uniform law of large numbers for β-mixing processes (Nobel and Dembo, 1993, Theorem 1). To estimate $\Omega_1(\tau, \gamma)$, we make use of a kernel method as described in Powell (1991) and Angrist et al. (2006):

$$\hat{\Omega}_1(\tau, \gamma) := \frac{1}{2nh_n}\sum_{t=1}^n I\{|y_t - \boldsymbol{x}'_t\tilde{\boldsymbol{\beta}}_{(1)}(\tau)| \le h_n\}\boldsymbol{z}_t(\gamma)\boldsymbol{z}_t(\gamma)', \tag{9}$$

where $\tilde{\boldsymbol{\beta}}_{(1)}(\tau)$ is any consistent estimator of $\boldsymbol{\beta}^*_{(1)}(\tau)$. We recommend to use

$$\tilde{\boldsymbol{\beta}}_{(1)}(\tau) := \arg\min_{\boldsymbol{b} \in \mathbb{R}^d} \frac{1}{n}\sum_{t=1}^n \rho_\tau(y_t - \boldsymbol{x}'_t\boldsymbol{b}). \tag{10}$$

The bandwidth h_n is chosen in such a way that $h_n \downarrow 0$ and $nh_n^2 \to \infty$ as $n \to \infty$. Then it is standard to show that under suitable regularity conditions, $\hat{\Omega}_1(\tau, \gamma)$ is uniformly consistent over $(\tau, \gamma) \in \mathcal{T} \times \Gamma$ by using the same argument as in Appendix A.1.4 of Angrist et al. (2006) coupled with some modifications. Therefore, we may estimate the matrix $V_{22}(\tau, \gamma)$ by

$$\hat{V}_{22}(\tau, \gamma) := \tau(1 - \tau)R\hat{\Omega}_1(\tau, \gamma)^{-1}\hat{\Omega}_0(\gamma, \gamma)\hat{\Omega}_1(\tau, \gamma)^{-1}R'.$$

For the sake of completeness, we provide a formal statement on the uniform consistency of $\hat{\Omega}_0(\gamma_1, \gamma_2)$ and $\hat{\Omega}_1(\tau, \gamma)$. A (sketch of) proof is given in Appendix C.

Lemma 2. *Assume conditions (C1)-(C6). Let $\tilde{\boldsymbol{\beta}}_{(1)}(\tau)$ be the estimator given by (10). Assume further that $\mathrm{E}[\|\boldsymbol{x}_t\|^{2p}] < \infty$ (p is given in condition (C2)), $h_n \downarrow 0$ and $nh_n^2 \to \infty$ as $n \to \infty$. Then $\hat{\Omega}_0(\gamma_1, \gamma_2) \xrightarrow{p} \Omega_0(\gamma_1, \gamma_2)$ uniformly over $(\gamma_1, \gamma_2) \in \Gamma \times \Gamma$, and $\hat{\Omega}_1(\tau, \gamma) \xrightarrow{p} \Omega_1(\tau, \gamma)$ uniformly over $(\tau, \gamma) \in \mathcal{T} \times \Gamma$.*

Hence, under the conditions of Lemma 2, $\hat{V}_{22}(\tau, \gamma)$ is uniformly consistent, so that the replacement of $V_{22}(\tau, \gamma)$ by $\hat{V}_{22}(\tau, \gamma)$ does not affect the distributional result of Corollary 1.

To compute approximate critical values of SW_n, we propose the following scheme. Take B as a large integer. For each $b = 1, \ldots, B$:

(i) Generate $\{u_t^b\}_{t=1}^n$ independent uniform random variables on $[0, 1]$.

(ii) Set $W_n^b(\tau, \gamma) := n^{-1/2} \sum_{t=1}^n \{\tau - I(u_t^b \leq \tau)\} \boldsymbol{z}_t(\gamma)$.

(iii) Compute the quantity

$$\widehat{SW}_n^b = \max_{(\tau, \gamma) \in \mathcal{T} \times \Gamma} W_n^b(\tau, \gamma)' \hat{\Omega}_1(\tau, \gamma)^{-1} R' \{\hat{V}_{22}(\tau, \gamma)\}^{-1} R \hat{\Omega}_1(\tau, \gamma)^{-1} W_n^b(\tau, \gamma).$$

Let $\hat{c}_{1-\alpha}^B$ denote the empirical $(1-\alpha)$-quantile of the simulated sample $\{\widehat{SW}_n^1, \ldots, \widehat{SW}_n^B\}$, where $\alpha \in (0, 1)$ is the nominal size. We reject the null hypothesis if SW_n is larger than $\hat{c}_{1-\alpha}^B$. In practice, the supremum in step (iii) is taken over a discretized subset of $\mathcal{T} \times \Gamma$.

The intuition behind this procedure is the fact that when the observations are independent, the first term on the Bahadur representation of the quantile regression estimator $\hat{\boldsymbol{\beta}}(\tau, \gamma)$:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \left[\tau - I\{y_t \leq \boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau)\} \right] \boldsymbol{z}_t(\gamma)$$

is conditionally pivotal given $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ under the null hypothesis. In fact, since $\boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau)$ is equal to the conditional τ-quantile of y_t given $\boldsymbol{z}_t$, the random variables $I\{y_1 \leq \boldsymbol{x}_1' \boldsymbol{\beta}_{(1)}^*(\tau)\}, \ldots, I\{y_n \leq \boldsymbol{x}_n' \boldsymbol{\beta}_{(1)}^*(\tau)\}$ are independent Bernoulli trials with success probability τ independent of $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ when the observations are independent. Under the present specification, the limiting null distribution of SW_n is the same as the one as if the observations were independent, so that it is expected that the method works also for dependent observations.

A formal justification of our simulation method is stated as follows. Let $\{u_t^*\}_{t=1}^n$ be independent uniform random variables on $[0, 1]$ independent of the sample $\{(y_t, \boldsymbol{z}_t')'\}_{t=1}^n$.

10

Define $W_n^*(\tau, \gamma) := n^{-1/2} \sum_{t=1}^n \{\tau - I(u_t^* \leq \tau)\} \boldsymbol{z}_t(\gamma)$ and

$$\widehat{SW}_n^* := \sup_{(\tau, \gamma) \in \mathcal{T} \times \Gamma} W_n^*(\tau, \gamma)' \hat{\Omega}_1(\tau, \gamma)^{-1} R' \{\hat{V}_{22}(\tau, \gamma)\}^{-1} R \hat{\Omega}_1(\tau, \gamma)^{-1} W_n^*(\tau, \gamma).$$

Let $\hat{c}_{1-\alpha}^*$ denote the $(1-\alpha)$-conditional quantile of $\widehat{SW}_n^*$ given the sample $\{(y_t, \boldsymbol{z}_t')'\}_{t=1}^n$, and let $c_{1-\alpha}$ denote the $(1-\alpha)$-quantile of the distribution given by (6) in Corollary 1.

Theorem 2. *Assume the conditions of Lemma 2. Let $\alpha \in (0, 1)$ be any constant. Let SW_n be the statistic (2) with $V_{22}(\tau, \gamma)$ replaced by $\hat{V}_{22}(\tau, \gamma)$. Then: (i) $\hat{c}_{1-\alpha}^* \xrightarrow{p} c_{1-\alpha}$; (ii) under H_0, $\mathrm{P}(SW_n > \hat{c}_{1-\alpha}^*) \to \alpha$; (iii) under H_1, $\mathrm{P}(SW_n > \hat{c}_{1-\alpha}^*) \to 1$.*

A proof of Theorem 2 is given in Appendix C.

Remark 2. Lee et al. (2011, Example 3.2) considered a simulation scheme very similar to ours to compute critical values of their test statistic. However, there are important differences between both methods: (i) the test statistic proposed by these authors is a likelihood ratio type test and not a Wald test as SW_n, (ii) Lee et al. (2011) did not allow for dependent observations, (iii) their test is pointwise, i.e., τ is fixed, and (iv) they did not provide a formal justification of their simulation scheme. Finally, the proof of Theorem 2 is different from that of Hansen (1996, Theorem 2) because we deal with a different statistic.

5   Simulation experiments

In this section, we conduct Monte Carlo simulations to evaluate the finite sample performance in terms of size and power of the proposed test. We are mainly interested in studying the properties of the supremum Wald test SW_n based on quantile regression over $\mathcal{T} \times \Gamma$. In addition, since the proposed test is applicable to pointwise testing, we also analyze the proposed test for fixed quantiles. Let $SW_n(\tau)$ denote the Wald test statistic based on quantile regression for fixed τ ($SW_n(\tau)$ corresponds to (2) with $\mathcal{T} = \{\tau\}$). For reference purposes, we compute Hansen's (1996) supremum Wald test SW_n^{LS} based on least squares.

In the experiments, we implement the tests of linearity, where we estimate a two-regime threshold model, fit quantile regression to compute SW_n and $SW_n(\tau)$ and least squares to compute SW_n^{LS}, and test the equality of the parameters in the two regimes. Therefore, SW_n and $SW_n(\tau)$ are testing linearity of the conditional quantile function against a two regime threshold model, while SW_n^{LS} tests linearity of the conditional mean function also against a two regime threshold model. We experiment with several different data generating processes, and for all different cases considered below, the

sample size n is equal to $n = 250, 500, 1000$ and the number of repetitions is $1,000$. The nominal sizes α under consideration are $\alpha = 0.01, 0.05$ and 0.10. When computing the approximate critical values based on the method described in the previous section, we take $B = 1,000$. The sets $\mathcal{T}$ and Γ are taken to be $\mathcal{T} = \{0.1, 0.15, \ldots, 0.9\}$ and $\Gamma = \{$ unconditional quantiles of y_t with quantile index ranging over $\{0.1, 0.2, \ldots, 0.9\}$ $\}$. To compute $\hat{\Omega}_1(\tau, \gamma)$ defined in (9), we use the bandwidth rule suggested in section 3.4.2 of Koenker (2005).

To investigate the empirical size of the test, we first consider a data generating process from a standard linear AR(1) model $y_t = \theta_0 + \theta_1 y_{t-1} + u_t$, with $\theta_0 = 0$, $\theta_1 = 0.5$ and $u_t \sim$ i.i.d. $U(-1, 1)$. In this model, $\boldsymbol{x}_t = (1, y_{t-1})'$, $q_t = y_{t-1}$ and $\mathcal{A}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$, and the data generating process satisfies the linear conditional quantile restriction : $Q_{y_t}(\tau | \mathcal{A}_{t-1}) = (\theta_0 + 2\tau - 1) + \theta_1 y_{t-1}$. The simulation results for empirical rejection rates are reported in Table 1. The SW_n test presents good empirical size. This statistic exhibits an empirical size closer to the nominal size than the least squares counterpart statistic, SW_n^{LS} for the three nominal sizes explored in the simulations. Interestingly, even for $n = 250$, the empirical sizes are very close to the nominal counterparts.

As for the pointwise tests based quantile regression, the results show fair size properties and variability across quantiles. Unsurprisingly, the SW_n test seems to provide more accurate estimates of the nominal size for $n = 1000$, especially for the upper quantiles of the distribution. Overall, these simulations show encouraging results and suggest that the tests developed here have empirical sizes close to their theoretical counterparts.

In order to evaluate the power of these tests we use two different sets of experiments. First, we assess the empirical power of the tests for three different two-regime models defined by different threshold nonlinearities (models 2, 3 and 4 below) in order to investigate advantages of quantile regression based tests versus least squares based tests. Second, we study the power of the test under local deviations from the linear null hypothesis. This family of models under the alternative hypothesis (model 5) is indexed by a local deviation parameter $n^{-1/2}c$.

For the first power experiment, we consider the following three models:

Model 2:

$$y_t = I(y_t > \gamma_0)(\theta_0 + \theta_1 y_{t-1} + u_t) + I(y_t \leq \gamma_0)(\theta_0 - \theta_1 y_{t-1} + u_t);$$

Model 3:

$$y_t = I(y_t > \gamma_0)(\theta_0 + \theta_1 y_{t-1} + (1 + \delta y_{t-1})u_t) + I(y_t \leq \gamma_0)(\theta_0 + \theta_1 y_{t-1} + (1 - \delta y_{t-1})u_t);$$

Model 4:

$$y_t = I(y_t > \gamma_0)(\theta_0 + \theta_1 y_{t-1} + (1 + \delta y_{t-1})u_t) + I(y_t \leq \gamma_0)(\theta_0 - \theta_1 y_{t-1} + (1 - \delta y_{t-1})u_t).$$

Note that these data generating processes satisfy the following restrictions:

Model 2:

$$Q_{y_t}(\tau | \mathcal{A}_{t-1}) = \begin{cases} (\theta_0 + 2\tau - 1) - \theta_1 y_{t-1}, & y_{t-1} \leq \gamma_0, \\ (\theta_0 + 2\tau - 1) + \theta_1 y_{t-1}, & y_{t-1} > \gamma_0; \end{cases}$$

$$\mathrm{E}[y_t | \mathcal{A}_{t-1}] = \begin{cases} \theta_0 - \theta_1 y_{t-1}, & y_{t-1} \leq \gamma_0, \\ \theta_0 + \theta_1 y_{t-1}, & y_{t-1} > \gamma_0; \end{cases}$$

Model 3:

$$Q_{y_t}(\tau | \mathcal{A}_{t-1}) = \begin{cases} (\theta_0 + 2\tau - 1) + (\theta_1 - \delta(2\tau - 1))y_{t-1}, & y_{t-1} \leq \gamma_0, \\ (\theta_0 + 2\tau - 1) + (\theta_1 + \delta(2\tau - 1))y_{t-1}, & y_{t-1} > \gamma_0; \end{cases}$$

$$\mathrm{E}[y_t | \mathcal{A}_{t-1}] = \begin{cases} \theta_0 + \theta_1 y_{t-1}, & y_{t-1} \leq \gamma_0, \\ \theta_0 + \theta_1 y_{t-1}, & y_{t-1} > \gamma_0; \end{cases}$$

Model 4:

$$Q_{y_t}(\tau | \mathcal{A}_{t-1}) = \begin{cases} (\theta_0 + 2\tau - 1) - (\theta_1 + \delta(2\tau - 1))y_{t-1}, & y_{t-1} \leq \gamma_0, \\ (\theta_0 + 2\tau - 1) + (\theta_1 + \delta(2\tau - 1))y_{t-1}, & y_{t-1} > \gamma_0; \end{cases}$$

$$\mathrm{E}[y_t | \mathcal{A}_{t-1}] = \begin{cases} \theta_0 - \theta_1 y_{t-1}, & y_{t-1} \leq \gamma_0, \\ \theta_0 + \theta_1 y_{t-1}, & y_{t-1} > \gamma_0. \end{cases}$$

Least squares based tests should be able to detect the presence of two regimes in models 2 and 4 where the threshold nonlinearity appears in the mean process, but not in model 3 where the mean process is identical in both regimes. Note that in models 3 and 4, the conditional variance process also exhibits a nonlinearity, that is, $\mathrm{Var}(y_t | \mathcal{A}_{t-1}, y_{t-1} \leq \gamma_0) \neq \mathrm{Var}(y_t | \mathcal{A}_{t-1}, y_{t-1} > \gamma_0)$. Quantile regression based tests, however, should detect two regimes in all three models because there are differences across regimes in the quantile process.

In all these cases we use $\gamma_0 = 0$, $\delta = 1$, $\theta_0 = 0$, $\theta_1 = 0.5$ and $u_t \sim$ i.i.d. $U(-1, 1)$. In these models, $\boldsymbol{x}_t = (1, y_{t-1})'$, $q_t = y_{t-1}$ and $\mathcal{A}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$.

The results for these three experiments are reported in Tables 2, 3 and 4, respectively. As expected, our proposed test SW_n has good power in all cases, and the power is increasing in n, which is consistent with the theoretical result. SW_n^{LS} also shows power for models 2 and 4, but this test lacks statistical power to reject the null hypothesis if the nonlinearities are not in the mean process, that is, as in model 3. In all cases we observe that for the pointwise quantile tests the empirical power increases as we move to the tails of the distributions. These results suggest that the uniform test is very powerful to reject linearity for models exhibiting threshold nonlinearities in either the conditional mean or conditional variance processes.

To complete the study of the power we perform a local power analysis:
Model 5:

$$y_t = I(y_t > \gamma_0)(\theta_0 + (\theta_1 + n^{-1/2}c)y_{t-1} + u_t) + I(y_t \le \gamma_0)(\theta_0 + (\theta_1 - n^{-1/2}c)y_{t-1} + u_t), \quad (11)$$

where $\gamma_0 = 0$, $c \in \{0, 1, 2, 3, 4, 5\}$, $\theta_0 = 0$, $\theta_1 = 0.5$, $u_t \sim$ i.i.d. $U(-1, 1)$ and $n = 250$.

The simulation results appear in Table 5. Note that for this case, where the threshold nonlinearity appears in the mean process, the least squares based test is expected to be more powerful than the developed SW_n test. In fact, the table shows that this is the case. Nevertheless, the simulation results confirm that the proposed SW_n test performs well in local deviations from the null hypothesis. Unreported simulations show that the power function is symmetric with respect to c.

6 Summary

We have proposed a uniform test for linearity against thresholds effects in the quantile regression framework. We have shown that the limiting null distribution of the proposed test is given by the supremum of a two-parameter chi-square process. This distribution depends on unknown parameters; hence critical values cannot be universally tabulated. We have developed a simulation method based on the Bahadur representation of the quantile regression estimator to approximate the critical values. The method makes the proposed test easy to implement. The Monte Carlo experiments show evidence that the proposed test has correct size and reasonable power against nonlinear threshold models.

Acknowledgments

A Proof of Theorem 1

In this section, we provide a proof of Theorem 1. The proof is based on a combination of modern empirical process techniques and the convexity technique developed in Kato (2009). Throughout the section, we assume all the conditions in Theorem 1.

A.1 Auxiliary result

For any fixed vector $\boldsymbol{v} \in \mathbb{R}^d$, define the stochastic processes

$$U_n(\tau, \gamma) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{v}'\boldsymbol{x}_t I(q_t \leq \gamma)[\tau - I\{y_t \leq \boldsymbol{x}_t'\boldsymbol{\beta}_{(1)}^*(\tau)\}],$$

$$V_n(\tau, \gamma, s) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\boldsymbol{v}'\boldsymbol{x}_t I(q_t \leq \gamma) I\{y_t \leq \boldsymbol{x}_t'(\boldsymbol{\beta}_{(1)}^*(\tau) + sn^{-1/2}\boldsymbol{v})\} \right.$$
$$\left. - \mathrm{E}[\boldsymbol{v}'\boldsymbol{x}_t I(q_t \leq \gamma) F(\boldsymbol{x}_t'(\boldsymbol{\beta}_{(1)}^*(\tau) + sn^{-1/2}\boldsymbol{v})|\boldsymbol{z}_t)] \right],$$

where $(\tau, \gamma, s) \in \mathcal{T} \times \Gamma \times [0, 1]$.

In this subsection, we study the asymptotic behaviors of U_n and V_n. Let $K(\tau, \lambda)$ denote a Kiefer process on $[0, 1] \times [0, \infty)$, i.e., $K(\tau, \lambda)$ is a zero-mean, continuous Gaussian process on $[0, 1] \times [0, \infty)$ with covariance kernel

$$\mathrm{E}[K(\tau_1, \lambda_1)K(\tau_2, \lambda_2)] = \lambda_1 \wedge \lambda_2(\tau_1 \wedge \tau_2 - \tau_1\tau_2).$$

We refer to Section 2.12 of van der Vaart and Wellner (1996) for Kiefer processes. Define $H(\gamma) := \mathrm{E}[(\boldsymbol{v}'\boldsymbol{x}_t)^2 I(q_t \leq \gamma)]$. Under conditions (C2) and (C3), the map $\gamma \mapsto H(\gamma)$ is continuous and non-decreasing.

Theorem 3. *Assume conditions (C1)-(C6). Then under H_0,*

(i) $U_n(\tau, \gamma) \Rightarrow K(\tau, H(\gamma))$ in $\ell^\infty(\mathcal{T} \times \Gamma)$ and

(ii) $V_n(\tau, \gamma, s) = V_n(\tau, \gamma, 0) + o_p(1)$ *uniformly over* $(\tau, \gamma, s) \in \mathcal{T} \times \Gamma \times [0, 1]$.

Proof. **Part (ii)**: We first prove Part (ii). For any compact subset $B \subset \mathbb{R}^d$, consider the class of functions

$$\mathcal{F} := \{f(y, q, \boldsymbol{x}) = \boldsymbol{v}'\boldsymbol{x}I(q \leq \gamma)I(y \leq \boldsymbol{x}'\boldsymbol{b}) : \gamma \in \Gamma, \boldsymbol{b} \in B\}.$$

We first show that $\mathcal{F}$ is a VC subgraph class. Indeed, let $g(y, q, \boldsymbol{x}) := \boldsymbol{v}'\boldsymbol{x}$, $\mathcal{F}_1 := \{f(y, q, \boldsymbol{x}) = I(q \leq \gamma) : \gamma \in \Gamma\}$ and $\mathcal{F}_2 := \{f(y, q, \boldsymbol{x}) = I(y \leq \boldsymbol{x}'\boldsymbol{b}) : \boldsymbol{b} \in B\}$. By Lemma 2.6.15 of van der Vaart and Wellner (1996), $\mathcal{F}_1$ and $\mathcal{F}_2$ are VC subgraph classes. Therefore, by Lemmas 2.6.18 (i) and (vi) of van der Vaart and Wellner (1996), $\mathcal{F} = (\mathcal{F}_1 \wedge \mathcal{F}_2) \cdot g$ is a VC-subgraph class.

Define the semimetric

$$\rho((\gamma_1, \boldsymbol{b}_1), (\gamma_2, \boldsymbol{b}_2)) := \left(\mathrm{E}[|\boldsymbol{v}'\boldsymbol{x}_t|^p \cdot |I(q_t \leq \gamma_1)I(y_t \leq \boldsymbol{x}_t'\boldsymbol{b}_1) - I(q_t \leq \gamma_2)I(y_t \leq \boldsymbol{x}_t'\boldsymbol{b}_2)|^p]\right)^{1/p},$$

where $p > 2$ is given in condition (C2). Because $\mathcal{F}$ is a VC subgraph class, by Theorem 2.1 of Arcones and Yu (1995) (more precisely their proof of Theorem 2.1), the stochastic process

$$(\gamma, \boldsymbol{b}) \in \Gamma \times B \mapsto \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{\boldsymbol{v}'\boldsymbol{x}_t I(q_t \leq \gamma)I(y_t \leq \boldsymbol{x}_t'\boldsymbol{b}) - \mathrm{E}[\boldsymbol{v}'\boldsymbol{x}_t I(q_t \leq \gamma)F(\boldsymbol{x}_t'\boldsymbol{b}|\boldsymbol{z}_t)]\}$$

$$=: G_n(\gamma, \boldsymbol{b}).$$

is stochastically ρ-equicontinuous over $\mathcal{T} \times \Gamma$, i.e., for any $\epsilon > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathrm{P}\left\{\sup_{[\delta]} |G_n(\gamma_1, \boldsymbol{b}_1) - G_n(\gamma_2, \boldsymbol{b}_2)| > \epsilon\right\} = 0,$$

where $[\delta] := \{((\gamma_1, \boldsymbol{b}_1), (\gamma_2, \boldsymbol{b}_2)) \in (\Gamma \times B)^2 : \rho((\gamma_1, \boldsymbol{b}_1), (\gamma_2, \boldsymbol{b}_2)) < \delta\}$.

Observe now that

$$\{\rho((\gamma_1, \boldsymbol{b}_1), (\gamma_2, \boldsymbol{b}_2))\}^p$$
$$\leq \mathrm{E}[|\boldsymbol{v}'\boldsymbol{x}_t|^p \{I(q_t \leq \gamma_1)I(y_t \leq \boldsymbol{x}_t'\boldsymbol{b}_1) - I(q_t \leq \gamma_2)I(y_t \leq \boldsymbol{x}_t'\boldsymbol{b}_2)\}^2]$$
$$\leq 2\mathrm{E}[|\boldsymbol{v}'\boldsymbol{x}_t|^p \{I(q_t \leq \gamma_1) - I(q_t \leq \gamma_2)\}^2] + 2\mathrm{E}[|\boldsymbol{v}'\boldsymbol{x}_t|^p \cdot |F(\boldsymbol{x}_t'\boldsymbol{b}_1|\boldsymbol{z}_t) - F(\boldsymbol{x}_t'\boldsymbol{b}_2|\boldsymbol{z}_t)|]$$
$$= 2|\mathrm{E}[|\boldsymbol{v}'\boldsymbol{x}_t|^p I(q_t \leq \gamma_1)] - \mathrm{E}[|\boldsymbol{v}'\boldsymbol{x}_t|^p I(q_t \leq \gamma_2)]| + 2\mathrm{E}[|\boldsymbol{v}'\boldsymbol{x}_t|^p \cdot |F(\boldsymbol{x}_t'\boldsymbol{b}_1|\boldsymbol{z}_t) - F(\boldsymbol{x}_t'\boldsymbol{b}_2|\boldsymbol{z}_t)|].$$

By conditions (C3) and (C4), we have

$$\sup_{(\tau,\gamma,s)\in\mathcal{T}\times\Gamma\times[0,1]} \rho((\gamma,\boldsymbol{\beta}^*_{(1)}(\tau) + n^{-1/2}s\boldsymbol{v}), (\gamma,\boldsymbol{\beta}^*_{(1)}(\tau))) = o(1).$$

Recall that $V_n(\tau,\gamma,s) = G_n(\gamma,\boldsymbol{\beta}^*_{(1)}(\tau) + n^{-1/2}s\boldsymbol{v})$. Thus, taking B sufficiently large (so that $\boldsymbol{\beta}^*_{(1)}(\tau) + n^{-1/2}s\boldsymbol{v} \in B$ for all $(\tau,s) \in \mathcal{T}\times[0,1]$), we obtain the first assertion because of the stochastic ρ-equicontinuity of the process G_n over $\Gamma\times B$.

Part (i): Observe that under H_0, $\{\tau - I\{y_t \le \boldsymbol{x}'_t\boldsymbol{\beta}^*_{(1)}(\tau)\}, t \in \mathbb{Z}\}$ is a martingale difference sequence with respect to $\{\mathcal{A}_{t-1}, t \in \mathbb{Z}\}$, i.e., $\mathrm{E}[I\{y_t \le \boldsymbol{x}'_t\boldsymbol{\beta}^*_{(1)}(\tau)\}|\mathcal{A}_{t-1}] = \tau$. The finite dimensional convergence follows from the martingale central limit theorem. It remains to show the stochastic equicontinuity of the process U_n. Decompose U_n as

$$U_n(\tau,\gamma) = \frac{\tau}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{v}'\boldsymbol{x}_t I(q_t \le \gamma) - \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{v}'\boldsymbol{x}_t I(q_t \le \gamma)I\{y_t \le \boldsymbol{x}'_t\boldsymbol{\beta}^*_{(1)}(\tau)\}$$

$$= \frac{\tau}{\sqrt{n}}\sum_{t=1}^n \{\boldsymbol{v}'\boldsymbol{x}_t I(q_t \le \gamma) - \mathrm{E}[\boldsymbol{v}'\boldsymbol{x}_t I(q_t \le \gamma)]\} - V_n(\tau,\gamma,0).$$

Define the stochastic process $\tilde{V}_n(\gamma) := n^{-1/2}\sum_{t=1}^n\{\boldsymbol{v}'\boldsymbol{x}_t I(q_t \le \gamma)\} - \mathrm{E}[\boldsymbol{v}'\boldsymbol{x}_t I(q_t \le \gamma)]\}$. Then $U_n(\tau,\gamma) = \tau\tilde{V}_n(\gamma) - V_n(\tau,\gamma,0)$. By the previous calculation, we see that

$$\rho((\gamma_1,\boldsymbol{\beta}^*_{(1)}(\tau_1)), (\gamma_2,\boldsymbol{\beta}^*_{(1)}(\tau_2))) \to 0, \text{ as } \|(\tau_1 - \tau_2, \gamma_1 - \gamma_2)\| \to 0,$$

which implies that the process $(\tau,\gamma) \mapsto V_n(\tau,\gamma,0)$ is stochastically equicontinuous over $\mathcal{T}\times\Gamma$ with respect to the Euclidean metric. Similarly, it is shown that the process $\gamma \mapsto \tilde{V}_n(\gamma)$ is stochastically equicontinuous over Γ with respect to the Euclidean metric. Therefore, by a standard argument, we obtain the desired conclusion. $\square$

A.2 Proof of Theorem 1

In this subsection, we provide a proof of Theorem 1. To this end, we introduce the local objective function

$$Z_n(\boldsymbol{u},\tau,\gamma) := \sum_{t=1}^n \left\{\rho_\tau\left(y_t - \boldsymbol{x}'_t\boldsymbol{\beta}^*_{(1)}(\tau) - n^{-1/2}\boldsymbol{u}'\boldsymbol{z}_t(\gamma)\right) - \rho_\tau\left(y_t - \boldsymbol{x}'_t\boldsymbol{\beta}^*_{(1)}(\tau)\right)\right\},$$

where $\boldsymbol{u} \in \mathbb{R}^{2d}$ and $(\tau,\gamma) \in \mathcal{T}\times\Gamma$. Observe that the normalized quantity $\sqrt{n}\{\hat{\boldsymbol{\beta}}(\tau,\gamma) - \boldsymbol{\beta}^*(\tau)\}$ minimizes $Z_n(\boldsymbol{u},\tau,\gamma)$ with respect to $\boldsymbol{u}$ for each fixed $(\tau,\gamma) \in \mathcal{T}\times\Gamma$.

To prove Theorem 1, we make use of Theorem 2 of Kato (2009), which makes the overall proof simple. Indeed, since $Z_n(\boldsymbol{u},\tau,\gamma)$ is now convex in $\boldsymbol{u}$, by Theorem 2 of

Kato (2009), it is sufficient to prove the following proposition.[2]

Proposition 2. *Assume conditions (C1)-(C6). Then under H_0,*

$$Z_n(\boldsymbol{u}, \tau, \gamma) = -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\tau - I\{y_t \leq \boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau)\} \right] \boldsymbol{u}' \boldsymbol{z}_t(\gamma) + \frac{1}{2} \boldsymbol{u}' \Omega_1(\tau, \gamma) \boldsymbol{u} + \Delta_n(\boldsymbol{u}, \tau, \gamma),$$

(12)

where $\sup_{(\tau,\gamma) \in \mathcal{T} \times \Gamma} |\Delta_n(\boldsymbol{u}, \tau, \gamma)| = o_p(1)$ for each fixed $\boldsymbol{u} \in \mathbb{R}^{2d}$, and

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\tau - I\{y_t \leq \boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau)\} \right] \boldsymbol{z}_t(\gamma) \Rightarrow W(\tau, \gamma) \ in \ (\ell^\infty(\mathcal{T} \times \Gamma))^{2d}.$$

(13)

Proof. Using Knight's (1998) identity

$$\rho_\tau(x - y) - \rho_\tau(x) = -y\{\tau - I(x \leq 0)\} + y \int_0^1 \{I(x \leq ys) - I(x \leq 0)\}ds,$$

we decompose $Z_n(\boldsymbol{u}, \tau, \gamma)$ into three parts:

$$Z_n(\boldsymbol{u}, \tau, \gamma)$$
$$= -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{u}' \boldsymbol{z}_t(\gamma) \left[\tau - I\{y_t \leq \boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau)\} \right]$$
$$+ \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{u}' \boldsymbol{z}_t(\gamma) \int_0^1 \sqrt{n}[I\{y_t - \boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau) \leq n^{-1/2} \boldsymbol{u}' \boldsymbol{z}_t(\gamma)s)\} - I\{y_t - \boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau) \leq 0\}]ds$$
$$=: Z_n^{(1)}(\boldsymbol{u}, \tau, \gamma) + Z_n^{(2)}(\boldsymbol{u}, \tau, \gamma)$$
$$= Z_n^{(1)}(\boldsymbol{u}, \tau, \gamma) + \mathrm{E}[Z_n^{(2)}(\boldsymbol{u}, \tau, \gamma)] + \{Z_n^{(2)}(\boldsymbol{u}, \tau, \gamma) - \mathrm{E}[Z_n^{(2)}(\boldsymbol{u}, \tau, \gamma)]\}$$
$$=: Z_n^{(1)}(\boldsymbol{u}, \tau, \gamma) + Z_n^{(21)}(\boldsymbol{u}, \tau, \gamma) + Z_n^{(22)}(\boldsymbol{u}, \tau, \gamma).$$

By Part (i) of Theorem 3, the weak convergence (13) follows. It remains to evaluate $Z_n^{(21)}(\boldsymbol{u}, \tau, \gamma)$ and $Z_n^{(22)}(\boldsymbol{u}, \tau, \gamma)$.

Evaluation of $Z_n^{(21)}(\boldsymbol{u}, \tau, \gamma)$: Observe that

$$Z_n^{(21)}(\boldsymbol{u}, \tau, \gamma) = \mathrm{E}\left[\boldsymbol{u}' \boldsymbol{z}_t(\gamma) \int_0^1 \sqrt{n}\{F(\boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau) + sn^{-1/2} \boldsymbol{u}' \boldsymbol{z}_t(\gamma))|\boldsymbol{z}_t) - \tau\}ds \right].$$

Since $F(\boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau)|\boldsymbol{z}_t) = \mathrm{E}[I\{y_t \leq \boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau)\}|\boldsymbol{z}_t] = \mathrm{E}[\mathrm{E}[I\{y_t \leq \boldsymbol{x}_t' \boldsymbol{\beta}_{(1)}^*(\tau)\}|\mathcal{A}_{t-1}]|\boldsymbol{z}_t] = \tau$

---

[2]In particular, Theorem 2 of Kato (2009) guarantees that there is no need to establish the uniform $n^{-1/2}$ rate for $\hat{\boldsymbol{\beta}}(\tau, \gamma)$ in the course of establishing the weak convergence of the process.

under H_0, we have

$$\sqrt{n}\{F(\boldsymbol{x}_t'\boldsymbol{\beta}_{(1)}^*(\tau) + sn^{-1/2}\boldsymbol{u}'\boldsymbol{z}_t(\gamma))|\boldsymbol{z}_t) - \tau\} - s\boldsymbol{u}'\boldsymbol{z}_t(\gamma)f(\boldsymbol{x}_t'\boldsymbol{\beta}_{(1)}^*(\tau)|\boldsymbol{z}_t)$$
$$= s\boldsymbol{u}'\boldsymbol{z}_t(\gamma)\int_0^1\{f(\boldsymbol{x}_t'\boldsymbol{\beta}_{(1)}^*(\tau) + svn^{-1/2}\boldsymbol{u}'\boldsymbol{z}_t(\gamma)|\boldsymbol{z}_t) - f(\boldsymbol{x}_t'\boldsymbol{\beta}_{(1)}^*(\tau)|\boldsymbol{z}_t)\}dv. \qquad (14)$$

Because of condition (C3), the right side of (14) goes to zero uniformly over $(\tau, \gamma, s) \in \mathcal{T} \times \Gamma \times [0, 1]$. Therefore, by the dominated convergence theorem, we have

$$\sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma}\left|Z_n^{(21)}(\boldsymbol{u}, \tau, \gamma) - \frac{1}{2}\boldsymbol{u}'\Omega_1(\tau, \gamma)\boldsymbol{u}\right| \to 0, \ \forall \boldsymbol{u} \in \mathbb{R}^{2d}.$$

<u>Evaluation of $Z_n^{(22)}(\boldsymbol{u}, \tau, \gamma)$</u>: Suppose for a moment that $\boldsymbol{u} \in \mathbb{R}^{2d}$ is arbitrarily fixed. Define

$$R_n(\tau, \gamma, s) := \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\boldsymbol{u}'\boldsymbol{z}_t(\gamma)I\{y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_{(1)}^*(\tau) \le sn^{-1/2}\boldsymbol{u}'\boldsymbol{z}_t(\gamma)\}\right.$$
$$\left. - \mathrm{E}[\boldsymbol{u}'\boldsymbol{z}_t(\gamma)F(\boldsymbol{x}_t'\boldsymbol{\beta}_{(1)}^*(\tau) + sn^{-1/2}\boldsymbol{u}'\boldsymbol{z}_t(\gamma)|\boldsymbol{z}_t)]\right],$$

where $(\tau, \gamma) \in \mathcal{T} \times \Gamma$ and $s \in \mathbb{R}$. We shall show that

$$\sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma}\sup_{s\in[0,1]}|R_n(\tau, \gamma, s) - R_n(\tau, \gamma, 0)| \xrightarrow{p} 0, \qquad (15)$$

which leads to $\sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma}|Z_n^{(22)}(\boldsymbol{u}, \tau, \gamma)| \xrightarrow{p} 0$. However, in view of

$$\boldsymbol{u}'\boldsymbol{z}_t(\gamma) = (\boldsymbol{u}_1 + \boldsymbol{u}_2)'\boldsymbol{x}_tI(q_t \le \gamma) + \boldsymbol{u}_1'\boldsymbol{x}_t\{1 - I(q_t \le \gamma)\},$$

(15) follows from Part (ii) of Theorem 3. $\qquad\qquad\square$

B   Conditions and proof for Proposition 1

B.1   Conditions for Proposition 1

We make the following conditions for Proposition 1.

(D1) For every $n \ge 1$, the process $\{(y_{n,t}, \boldsymbol{z}_{n,t}')', t \in \mathbb{Z}\}$ is strict stationary and β-mixing with β-mixing coefficients $\beta_n(j)$. Moreover, there exists a sequence of non-increasing constants $\beta(j)$ such that $\sup_{n\ge1}\beta_n(j) \le \beta(j)$ and $\beta(j) = O(j^{-l})$ with $l > p/(p-2)$, where $p > 2$ is given in condition (D2) below.

(D2) $\sup_{n\ge1}\mathrm{E}[\|\boldsymbol{x}_{n,t}\|^p] < \infty$ for some $p > 2$.

19

(D3) Let $F_n(\cdot|\boldsymbol{z})$ denote the conditional distribution function of $y_{n,t}$ given $\boldsymbol{z}_{n,t} = \boldsymbol{z}$. Assume that $F_n(\cdot|\boldsymbol{z})$ has a Lebesgue density $f_n(\cdot|\boldsymbol{z})$ such that

(i) $|f_n(y|\boldsymbol{z})| \leq C_f$, and

(ii) $|f_n(y_1|\boldsymbol{z}) - f(y_2|\boldsymbol{z})| \leq C_f|y_1 - y_2|$,

where $C_f > 0$ is a constant independent of n.

(D4) The distribution functions of $q_{n,t}$ are continuous uniformly in n.

(D5) Define the matrices

$$\Omega_{n,0}(\gamma_1, \gamma_2) := \mathrm{E}[\boldsymbol{z}_{n,t}(\gamma_1)\boldsymbol{z}_{n,t}(\gamma_2)'],$$

$$\Omega_{n,2}(\tau, \gamma_1, \gamma_2) := \mathrm{E}[f_n(\boldsymbol{x}_{n,t}'\boldsymbol{\beta}_{(1)}^*(\tau)|\boldsymbol{z}_{n,t})\boldsymbol{z}_{n,t}(\gamma)\boldsymbol{z}_{n,t}(\gamma)'], \quad \Omega_{n,1}(\tau, \gamma) := \Omega_{n,2}(\tau, \gamma, \gamma).$$

There exist $d \times d$ matrices $\Omega_0(\gamma_1, \gamma_2)$ and $\Omega_2(\tau, \gamma_1, \gamma_2)$ such that $\Omega_{n,0}(\gamma_1, \gamma_2) \to \Omega_0(\gamma_1, \gamma_2)$ and $\Omega_{n,2}(\tau, \gamma_1, \gamma_2) \to \Omega_2(\tau, \gamma_1, \gamma_2)$ uniformly over $(\gamma_1, \gamma_2) \in \Gamma \times \Gamma$ and $(\tau, \gamma_1, \gamma_2) \in \mathcal{T} \times \Gamma \times \Gamma$, respectively. Moreover, $\Omega_0(\gamma_1, \gamma_2)$ and $\Omega_1(\tau, \gamma) := \Omega_2(\tau, \gamma, \gamma)$ are positive definite uniformly over $(\gamma_1, \gamma_2) \in \Gamma \times \Gamma$ and $(\tau, \gamma) \in \mathcal{T} \times \Gamma$, respectively.

B.2   Proof of Proposition 1

As in the proof of Theorem 1, define

$$Z_n(\boldsymbol{u}, \tau, \gamma) := \sum_{t=1}^n \left\{ \rho_\tau\left(y_{n,t} - \boldsymbol{x}_{n,t}'\boldsymbol{\beta}_{(1)}^*(\tau) - n^{-1/2}\boldsymbol{u}'\boldsymbol{z}_{n,t}(\gamma)\right) - \rho_\tau\left(y_{n,t} - \boldsymbol{x}_{n,t}'\boldsymbol{\beta}_{(1)}^*(\tau)\right) \right\},$$

where $\boldsymbol{u} \in \mathbb{R}^{2d}$ and $(\tau, \gamma) \in \mathcal{T} \times \Gamma$. An inspection of the proof of Proposition 2 shows that under H_{1n}, we still have the expansion

$$Z_n(\boldsymbol{u}, \tau, \gamma) = -\frac{1}{\sqrt{n}}\sum_{t=1}^n \left[\tau - I\{y_{n,t} \leq \boldsymbol{x}_{n,t}'\boldsymbol{\beta}_{(1)}^*(\tau)\}\right]\boldsymbol{u}'\boldsymbol{z}_{n,t}(\gamma) + \frac{1}{2}\boldsymbol{u}'\Omega_1(\tau, \gamma)\boldsymbol{u} + \Delta_n(\boldsymbol{u}, \tau, \gamma),$$

where $\sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma}|\Delta_n(\boldsymbol{u}, \tau, \gamma)| = o_p(1)$ for each fixed $\boldsymbol{u} \in \mathbb{R}^{2d}$. Here the difference arises since $\tau - I\{y_{n,t} \leq \boldsymbol{x}_{n,t}'\boldsymbol{\beta}_{(1)}^*(\tau)\}$ are not centered. Fix $\boldsymbol{u} \in \mathbb{R}^{2d}$, and define

$$Z_n^{(1)}(\tau, \gamma) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \left[\tau - I\{y_{n,t} \leq \boldsymbol{x}_{n,t}'\boldsymbol{\beta}_{(1)}^*(\tau)\}\right]\boldsymbol{u}'\boldsymbol{z}_{n,t}(\gamma).$$

Let us write $\boldsymbol{\beta}_n^*(\tau) = (\boldsymbol{\beta}_{(1)}^*(\tau)', \boldsymbol{\beta}_{(2),n}(\tau)')'$ where $\boldsymbol{\beta}_{(2),n}(\tau) = n^{-1/2}\boldsymbol{c}(\tau)$. Observe that

$$Z_n^{(1)}(\tau, \gamma) = \frac{1}{\sqrt{n}} \sum_{t=1}^n [\tau - I\{y_{n,t} \le \boldsymbol{z}_{n,t}(\gamma_0)'\boldsymbol{\beta}_n^*(\tau)\}]\, \boldsymbol{u}'\boldsymbol{z}_{n,t}(\gamma)$$

$$+ \frac{1}{\sqrt{n}} \sum_{t=1}^n \left[I\{y_{n,t} \le \boldsymbol{z}_{n,t}(\gamma_0)'\boldsymbol{\beta}_n^*(\tau)\} - I\{y_{n,t} \le \boldsymbol{x}_{n,t}'\boldsymbol{\beta}_{(1)}^*(\tau)\} \right] \boldsymbol{u}'\boldsymbol{z}_{n,t}(\gamma)$$

$$=: Z_n^{(11)}(\tau, \gamma) + Z_n^{(12)}(\tau, \gamma)$$

$$= Z_n^{(11)}(\tau, \gamma) + \mathrm{E}[Z_n^{(12)}(\tau, \gamma)] + Z_n^{(12)}(\tau, \gamma) - \mathrm{E}[Z_n^{(12)}(\tau, \gamma)]$$

$$=: Z_n^{(11)}(\tau, \gamma) + \mathrm{E}[Z_n^{(12)}(\tau, \gamma)] + Z_n^{(13)}(\tau, \gamma).$$

By Theorem 3 (with suitable modifications), we can show that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \left[\tau - I\{y_{n,t} \le \boldsymbol{z}_{n,t}(\gamma_0)'\boldsymbol{\beta}_{(n)}^*(\tau)\} \right] \boldsymbol{z}_{n,t}(\gamma) \Rightarrow W(\tau, \gamma), \text{ in } (\ell^\infty(\mathcal{T} \times \Gamma))^{2d},$$

where $W(\tau, \gamma)$ is as in Theorem 1 (with $\Omega_0(\gamma_1, \gamma_2)$ replaced by (7)), and

$$\sup_{(\tau, \gamma) \in \mathcal{T} \times \Gamma} |Z_n^{(13)}(\tau, \gamma)| = o_p(1).$$

By using Taylor's theorem, we also have

$$\mathrm{E}[Z_n^{(12)}(\tau, \gamma)] = \boldsymbol{u}'\mathrm{E}[f_n(\boldsymbol{x}_{n,t}'\boldsymbol{\beta}_{(1)}(\tau)|\boldsymbol{z}_{n,t})\boldsymbol{z}_{n,t}(\gamma)\boldsymbol{z}_{n,t}(\gamma_0)']R'\boldsymbol{c}(\tau) + o(1)$$

$$= \boldsymbol{u}'\Omega_2(\tau, \gamma, \gamma_0)R'\boldsymbol{c}(\tau) + o(1).$$

uniformly over $(\tau, \gamma) \in \mathcal{T} \times \Gamma$. Hence by Theorem 2 of Kato (2009), we conclude that

$$\sqrt{n}\{\hat{\boldsymbol{\beta}}(\tau, \gamma) - (\boldsymbol{\beta}_{(1)}^*(\tau)', \boldsymbol{0}')')\} \Rightarrow \Omega_1(\tau, \gamma)^{-1}(W(\tau, \gamma) + \Omega_2(\tau, \gamma, \gamma_0)R'\boldsymbol{c}(\tau)), \text{ in } (\ell^\infty(\mathcal{T} \times \Gamma))^{2d},$$

which implies the conclusion of the proposition. $\square$

C Proofs for Section 4

C.1 Proof of Lemma 2

We only provide a sketch of the proof. It is standard to show that when the observations are independent and identically distributed, $\hat{\Omega}_0(\gamma_1, \gamma_2) \to \Omega_0(\gamma_1, \gamma_2)$ almost surely uniformly over $(\gamma_1, \gamma_2) \in \Gamma \times \Gamma$. By Theorem 1 of Nobel and Dembo (1993), the uniform consistency of $\hat{\Omega}_0(\gamma_1, \gamma_2)$ now follows under the present conditions. By Appendix A.1.1 of Angrist et al. (2006) together with the same theo-

rem of Nobel and Dembo (1993), it is shown that $\tilde{\beta}_1(\tau)$ is uniformly consistent: $\sup_{\tau \in \mathcal{T}} \|\tilde{\boldsymbol{\beta}}_{(1)}(\tau) - \boldsymbol{\beta}^*_{(1)}(\tau)\| \overset{p}{\to} 0$. The proof of the uniform consistency of $\hat{\Omega}_1(\tau, \gamma)$ is similar to the proof given in Appendix A.1.4 of Angrist et al. (2006). Fix any $1 \le j, k \le d$. For any compact subset $B \subset \mathbb{R}^d$ and positive constant H, consider the class of functions $\mathcal{F} := \{f(y, q, \boldsymbol{x}) = I(|y - \boldsymbol{x}'\boldsymbol{b}| \le h)I(q \le \gamma)x_j x_k : \gamma \in \Gamma, \boldsymbol{b} \in B, h \in (0, H]\}$, where $\boldsymbol{x} = (x_1, \ldots, x_d)'$. As in the proof of Theorem 3, it is shown that the class $\mathcal{F}$ is a VC subgraph class. Using the uniform central limit theorem for β-mixing processes (Arcones and Yu, 1995, Theorem 2.1), and following the proof given in Appendix A.1.4 of Angrist et al. (2006), it is shown that, under the present conditions,

$$\frac{1}{2nh_n} \sum_{t=1}^{n} I(|y_t - \boldsymbol{x}_t' \tilde{\boldsymbol{\beta}}_{(1)}(\tau)| \le h_n)I(q_t \le \gamma)x_{tj}x_{tk}$$

$$= \frac{1}{2h_n}\mathrm{E}\left[I(|y_t - \boldsymbol{x}_t'\boldsymbol{b}| \le h_n)I(q_t \le \gamma)x_{tj}x_{tk}\right]|_{\boldsymbol{b}=\tilde{\boldsymbol{\beta}}_{(1)}(\tau)} + O_p(n^{-1/2}h_n^{-1})$$

$$= \mathrm{E}[f(\boldsymbol{x}_t'\boldsymbol{\beta}^*_{(1)}(\tau)|\boldsymbol{z}_t)I(q_t \le \gamma)x_{tj}x_{tk}] + o_p(1),$$

and the convergence is uniform over $(\tau, \gamma) \in \mathcal{T} \times \Gamma$. With the same proof, it is shown that, uniformly over $\tau \in \mathcal{T}$,

$$\frac{1}{2nh_n} \sum_{t=1}^{n} I(|y_t - \boldsymbol{x}_t'\tilde{\boldsymbol{\beta}}_{(1)}(\tau)| \le h_n)x_{tj}x_{tk} \overset{p}{\to} \mathrm{E}[f(\boldsymbol{x}_t'\boldsymbol{\beta}^*_{(1)}(\tau)|\boldsymbol{z}_t)x_{tj}x_{tk}].$$

Therefore, we have $\hat{\Omega}_1(\tau, \gamma) \overset{p}{\to} \Omega_1(\tau, \gamma)$ uniformly over $(\tau, \gamma) \in \mathcal{T} \times \Gamma$.

C.2   Proof of Theorem 2

Let SW denote a random variable having the same distribution as (6) in Corollary 1. We first point out that SW has an absolutely continuous distribution with bounded density by Proposition 11.4 of Davidov et al. (1998). Therefore, by a routine argument, (ii) and (iii) follow from (i). For example, to see that (ii) follows from (i), we first observe that (i) implies that there exists a sequence $\epsilon_n \to 0$ such that $\mathrm{P}(|\hat{c}^*_{1-\alpha} - c_{1-\alpha}| \le$

$\epsilon_n) \to 1$, so that

$$
\begin{aligned}
&\mathrm{P}(SW_n > \hat{c}^*_{1-\alpha}) \\
&= \mathrm{P}(\{SW_n > \hat{c}^*_{1-\alpha}\} \cap \{\hat{c}^*_{1-\alpha} \geq c_{1-\alpha} - \epsilon_n\}) + o(1) \\
&\leq \mathrm{P}(SW_n > c_{1-\alpha} - \epsilon_n) + o(1) \\
&\leq \mathrm{P}(SW > c_{1-\alpha} - \epsilon_n) + [\mathrm{P}(SW_n > c_{1-\alpha} - \epsilon_n) - \mathrm{P}(SW > c_{1-\alpha} - \epsilon_n)] + o(1) \\
&\leq \mathrm{P}(SW > c_{1-\alpha} - \epsilon_n) + \sup_{c \in \mathbb{R}} |\mathrm{P}(SW_n > c) - \mathrm{P}(SW > c)| + o(1) \\
&= \mathrm{P}(SW > c_{1-\alpha} - \epsilon_n) + o(1),
\end{aligned}
$$

where the last equality follows from the fact that $SW_n \Rightarrow SW$ under H_0, SW has a continuous distribution and Polya's theorem. Because SW has a bounded density function, we have $\mathrm{P}(SW > c_{1-\alpha} - \epsilon_n) = \mathrm{P}(SW > c_{1-\alpha}) + o(1) = \alpha + o(1)$, so that $\mathrm{P}(SW_n > \hat{c}^*_{1-\alpha}) \leq \alpha + o(1)$. A similar argument leads to that $\mathrm{P}(SW_n > \hat{c}^*_{1-\alpha}) \geq \alpha - o(1)$, which implies (i). By an analogous argument, it is shown that (iii) follows from (i).

In what follows, we wish to show (i). To this end, we need the concept of "conditional weak convergence in probability". We follow the definition given in van der Vaart and Wellner (1996, p.181). We wish to show that, given $\{(y_t, z'_t)'\}_{i=1}^n$, $\widehat{SW}_n^*$ converges weakly to SW in probability. Once this is established, the assertion of (i) follows because of the fact that the quantile function of SW is everywhere continuous by Theorem 11.1 (iv) of Davidov et al. (1998) (recall that the distribution function of SW is continuous). By a standard argument, it suffices to show that, given $\{z_t\}_{t=1}^n$, in the space $(\ell^\infty(\mathcal{T} \times \Gamma))^{2d}$, W_n^* converges weakly in probability to the Gaussian process W defined in Theorem 1.

We adapt the proof of van der Vaart and Wellner (1996, Theorem 2.9.6) to the present context (some modifications are required because the situation is different from that theorem; in particular, $z_1, \dots, z_n$ are now dependent and the quantile index τ is involved in the simulated term $I(u_t^* \leq \tau)$, so their Lemma 2.9.1 can not be used in the present case). Introduce any norm $\|\cdot\|_{(\ell^\infty(\mathcal{T} \times \Gamma))^{2d}}$ on $(\ell^\infty(\mathcal{T} \times \Gamma))^{2d}$ that induces the topology of $(\ell^\infty(\mathcal{T} \times \Gamma))^{2d}$ (for example, take $\|f\|_{(\ell^\infty(\mathcal{T} \times \Gamma))^{2d}} = \sup_{(\tau,\gamma) \in \mathcal{T} \times \Gamma} \|f(\tau, \gamma)\|$ for $f \in (\ell^\infty(\mathcal{T} \times \Gamma))^{2d}$). Let BL_1 denote the space of all Lipschitz continuous functions $(\ell^\infty(\mathcal{T} \times \Gamma))^{2d} \to [-1, 1]$ with Lipschitz constant bounded by 1. We have to show that

$$
\sup_{h \in BL_1} |\mathrm{E}[h(W_n^*)|z_1^n] - \mathrm{E}[h(W)]| \xrightarrow{p} 0,
$$

where $z_1^n := \{z_t\}_{t=1}^n$. We divide the proof into three steps.

Step 1: Because $\mathcal{T} \times \Gamma$ is compact in $\mathbb{R}^2$, for any $\delta > 0$, there exists a finite set $\{(\tau_k, \gamma_k)\}_{k=1}^K$ such that $\sup_{(\tau,\gamma) \in \mathcal{T} \times \Gamma} \min_{1 \leq k \leq K} \|(\tau,\gamma) - (\tau_k, \gamma_k)\| \leq \delta$. Let Π_δ denote the map from $\mathcal{T} \times \Gamma$ to $\{(\tau_k, \gamma_k)\}_{k=1}^K$ such that $\|(\tau,\gamma) - \Pi_\delta(\tau,\gamma)\| \leq \delta$ for all $(\tau,\gamma) \in \mathcal{T} \times \Gamma$. Define the stochastic process $W \circ \Pi_\delta$ on $\mathcal{T} \times \Gamma$ by $W \circ \Pi_\delta(\tau,\gamma) := W(\Pi_\delta(\tau,\gamma))$. Then because W is uniformly continuous on $\mathcal{T} \times \Gamma$, as $\delta \downarrow 0$, $W \circ \Pi_\delta \to W$ in $(\ell^\infty(\mathcal{T} \times \Gamma))^{2d}$. Therefore, by the dominated convergence theorem, we have

$$\lim_{\delta \downarrow 0} \sup_{h \in BL_1} |\mathrm{E}[h(W \circ \Pi_\delta)] - \mathrm{E}[h(W)]| = 0. \tag{16}$$

Step 2: Fix any $\delta > 0$. We wish to show that, with $W_n^* \circ \Pi_\delta(\tau,\gamma) := W_n^*(\Pi_\delta(\tau,\gamma))$,

$$\sup_{h \in BL_1} |\mathrm{E}[h(W_n^* \circ \Pi_\delta)|\boldsymbol{z}_1^n] - \mathrm{E}[h(W \circ \Pi_\delta)]| \overset{p}{\to} 0. \tag{17}$$

To this end, it suffices to show that conditionally on $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots$,

$$(W_n^*(\tau_k, \gamma_k))_{1 \leq k \leq K} \overset{d}{\to} (W(\tau_k, \gamma_k))_{1 \leq k \leq K}, \tag{18}$$

for almost every sequence $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots$. Because of the ergodicity, we have: (a) $\hat{\Omega}_0(\gamma_k, \gamma_l) = n^{-1} \sum_{t=1}^n \boldsymbol{z}_t(\gamma_k)\boldsymbol{z}_t(\gamma_l)' \to \Omega_0(\gamma_k, \gamma_l)$ for all $1 \leq k, l \leq K$ almost surely, and (b) $\max_{1 \leq t \leq n} \|\boldsymbol{x}_t\|/\sqrt{n} \to 0$ almost surely. On the intersection of these two events (events (a) and (b)), it is shown by the Lindberg-Feller central limit theorem that (18) holds conditionally on $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots$. Therefore, we obtain the desired claim.

Step 3: We wish to show that

$$\lim_{\delta \downarrow 0} \lim_{n \to \infty} \mathrm{E}\left[\sup_{h \in BL_1} |\mathrm{E}[h(W_n^*)|\boldsymbol{z}_1^n] - \mathrm{E}[h(W_n^* \circ \Pi_\delta)|\boldsymbol{z}_1^n]|\right] = 0. \tag{19}$$

Observe that $|\mathrm{E}[h(W_n^*)|\boldsymbol{z}_1^n] - \mathrm{E}[h(W_n^* \circ \Pi_\delta)|\boldsymbol{z}_1^n]| \leq \mathrm{E}[|h(W_n^*) - h(W_n^* \circ \Pi_\delta)||\boldsymbol{z}_1^n]$, so that

$$\mathrm{E}\left[\sup_{h \in BL_1} |\mathrm{E}[h(W_n^*)|\boldsymbol{z}_1^n] - \mathrm{E}[h(W_n^* \circ \Pi_\delta)|\boldsymbol{z}_1^n]|\right] \leq \mathrm{E}\left[\sup_{h \in BL_1} |h(W_n^*) - h(W_n^* \circ \Pi_\delta)|\right].$$

By essentially the same proof as that of Theorem 3, it is shown that $W_n^* \Rightarrow W$ in $(\ell^\infty(\mathcal{T} \times \Gamma))^{2d}$ unconditionally. Thus, by Theorem 1.5.7 of van der Vaart and Wellner (1996), for any $\epsilon > 0$, there exist a constant $\delta = \delta_\epsilon > 0$ and a positive integer n_ϵ such that for all $n \geq n_\epsilon$,

$$\mathrm{P}\left\{\|W_n^* - W_n^* \circ \Pi_\delta\|_{(\ell^\infty(\mathcal{T} \times \Gamma))^{2d}} > \epsilon\right\} \leq \epsilon.$$

Assume that $n \geq n_\epsilon$. Let A_0 denote the event that $\|W_n^* - W_n^* \circ \Pi_\delta\|_{(\ell^\infty(\mathcal{T} \times \Gamma))^{2d}} \leq \epsilon$. On

the event A_0, $|h(W_n^*) - h(W_n^* \circ \Pi_\delta)| \leq \|W_n^* - W_n^* \circ \Pi_\delta\|_{(\ell^\infty(\mathcal{T} \times \Gamma))^{2d}} \leq \epsilon$, while on the event A_0^c, $|h(W_n^*) - h(W_n^* \circ \Pi_\delta)| \leq 2$. Observe that $P(A_0^c) \leq \epsilon$. Therefore, we have

$$\mathrm{E}\left[\sup_{h \in BL_1} |h(W_n^*) - h(W_n^* \circ \Pi_\delta)|\right] \leq 3\epsilon.$$

This implies (19).

Finally, for any $\delta > 0$,

$$\sup_{h \in BL_1} |\mathrm{E}[h(W_n^*)|\boldsymbol{z}_1^n] - \mathrm{E}[h(W)]| \leq \sup_{h \in BL_1} |\mathrm{E}[h(W_n^*)|\boldsymbol{z}_1^n] - \mathrm{E}[h(W_n^* \circ \Pi_\delta)|\boldsymbol{z}_1^n]|$$
$$+ \sup_{h \in BL_1} |\mathrm{E}[h(W_n^* \circ \Pi_\delta)|\boldsymbol{z}_1^n] - \mathrm{E}[h(W \circ \Pi_\delta)]|$$
$$+ \sup_{h \in BL_1} |\mathrm{E}[h(W \circ \Pi_\delta)] - \mathrm{E}[h(W)]|.$$

The expectation of the first term on the right side goes to zero as letting $n \to \infty$ first and then $\delta \downarrow 0$ by (19); the expectation of the second term goes to zero as $n \to \infty$ by (17) and the dominated convergence theorem; and the last term goes to zero as $\delta \downarrow 0$ by (16). Therefore, we have

$$\mathrm{E}\left[\sup_{h \in BL_1} |\mathrm{E}[h(W_n^*)|\boldsymbol{z}_1^n] - \mathrm{E}[h(W)]|\right] \to 0.$$

By Markov's inequality, we obtain the desired conclusion.

$\square$

References

Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817-858.

Andrews, D.W.K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62** 1383-1414.

Angrist, J., Chernozhukov, V. and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the US wage structure. *Econometrica* **74** 539-563.

Arcones, M. and Yu, B. (1995). Central limit theorems for empirical and *U*-processes of stationary mixing sequences. *J. Theoret. Probab.* **7** 47-71.

Cai, Y. (2010). Forecasting for quantile self-exciting threshold autoregressive time series models. *Biometrika* **97** 199-208.

Cai, Y. and Stander, J. (2008). Quantile self-exciting threshold time series models. *J. Time Series Anal.* **29** 187-202.

Caner, M. (2002). A note on least absolute deviation estimation of a threshold model. *Econometric Theory* **18** 800-814.

Carrasco, M. (2002). Misspecified structural change, threshold, and Markov-switching models. *J. Econometrics* **109** 239-273.

Chan, K.S. (1990). Testing for threshold autoregression. *Ann. Statist.* **18** 1886-1894.

Chernozhukov, V. and Fernández-Val, I. (2005). Subsampling inference on quantile regression processes. *Sankhya* **67** 253-276.

Davidov, Y., Lifshits, M. and Smorodina, N. (1998). *Local Properties of Distributions of Stochastic Functions* (Transaction of Mathematical Monographs, Vol. 173). American Mathematical Society.

Davies, R. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** 247-254.

Davies, R. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33-43.

Escanciano, J.C. and Velasco, C. (2010). Specification tests of parametric dynamic conditional quantiles. *J. Econometrics* **159** 209-221.

Fan, J. and Yao, Q. (2005). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag.

Gutenbrunner, C. and Jurečková, J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20** 305-330.

Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64** 413-430.

Hansen, B.E. (1997). Inference in TAR models. *Stud. Nonlinear Dyn. Econom.* **2** 1-14.

Hansen, B.E. (2000). Sample splitting and threshold regression. *Econometrica* **68** 575-603.

He, X. and Zhu, L.X. (2003). A lack-of-fit test for quantile regression. *J. Amer. Stat. Assoc.* **98** 1013-1022.

Horowitz, J.L. and Spokoiny, V.G. (2002). An adaptive, rate-optimal test of linearity of median regression models. *J. Amer. Stat. Assoc.* **97** 822-835.

Kato, K. (2009). Asyptotics for argmin processes: convexity arguments. *J. Multivariate Anal.* **100** 1816-1829.

Knight, K. (1998). Limiting distribution for L_1 regression estimators under general conditions. *Ann. Statist.* **26** 755-770.

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

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46** 33-50.

Koenker, R. and Machado, J.A.F. (1999). Goodness-of-fit and related inference processes for quantile regression. *J. Amer. Stat. Assoc.* **94** 1296-1310.

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

Komunjer, I. and Vuong, Q. (2010a). Semiparametric efficiency bound in time-series models for conditional quantiles. *Econometric Theory* **26** 383-405.

Komunjer, I. and Vuong, Q. (2010b). Efficient estimation in dynamic conditional quantile models. *J. Econometrics* **157** 272-285.

Lee, S., Seo, M.H. and Shin, Y. (2011). Testing for threshold effects in regression models. *J. Amer. Stat. Assoc.* **106** 220-231.

Nobel, A. and Dembo, A. (1993). A note on uniform laws of averages for dependent processes. *Statist. Probab. Lett.* **17** 169-172.

Otsu, T. (2008). Conditional empirical likelihood estimation and inference for quantile regression. *J. Econometrics* **142** 508-538.

Powell, J.L. (1991). Estimation of monotonic regression models under quantile restrictions. In: *Nonparametric and Semiparametric Models in Econometrics* (eds. W. Barnett, J.L. Powell and G. Tauchen). Cambridge University Press.

Qu, Z. (2008). Testing for structural change in quantile regression. *J. Econometrics* **146** 170-184.

Su, L. and Xiao, Z. (2008). Testing for parameter stability in quantile regression models. *Statist. Probab. Lett.* **78** 2768-2775.

Su, L. and White, H. (2012). Conditional independence specification testing for dependent processes with local polynomial quantile regression. *Advances in Econometrics* **29** 355-434.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis* (Lecture Notes in Statistics). Springer-Verlag.

Tong, H. and Lim, K. (1980). Threshold autoregression, limit cycles and cyclical data. *J. Roy. Stat. Soc. Ser. B Stat. Methodol.* **4** 245-292.

Tsay, R. (1989). Testing and modeling threshold autoregressive processes. *J. Amer. Stat. Assoc.* **18** 931-955.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag.

Whang, Y.-J. (2005). Consistent specification testing for quantile regression models. In: *Frontiers of Analysis and Applied Research: Essays in Honor of Peter C.B. Phillips* (eds. D. Corbae, S.N. Durlauf, and B.E. Hansen). Cambridge University Press.

Zheng, J.X. (1998). A consistent nonparametric test of parametric regression model under conditional quantile restrictions. *Econometric Theory* **14** 123-138.

Tables

Table 1: Size - One-regime AR(1) model

	$n = 250$			$n = 500$			$n = 1000$		
	Theoretical size								
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
0.10	0.058	0.135	0.204	0.034	0.109	0.154	0.037	0.099	0.173
0.15	0.033	0.074	0.106	0.011	0.027	0.042	0.006	0.021	0.036
0.20	0.021	0.059	0.091	0.012	0.037	0.066	0.007	0.036	0.065
0.25	0.026	0.057	0.096	0.009	0.049	0.095	0.019	0.055	0.106
0.30	0.024	0.072	0.119	0.016	0.060	0.111	0.025	0.077	0.157
0.35	0.025	0.086	0.139	0.026	0.069	0.137	0.033	0.090	0.160
0.40	0.019	0.078	0.147	0.036	0.094	0.160	0.029	0.094	0.172
0.45	0.027	0.092	0.158	0.033	0.104	0.173	0.032	0.088	0.154
0.50	0.031	0.112	0.174	0.045	0.111	0.182	0.032	0.097	0.169
0.55	0.030	0.094	0.173	0.046	0.105	0.172	0.030	0.093	0.157
0.60	0.025	0.102	0.150	0.044	0.098	0.177	0.037	0.097	0.153
0.65	0.020	0.079	0.144	0.036	0.107	0.177	0.034	0.109	0.163
0.70	0.020	0.078	0.135	0.021	0.094	0.160	0.041	0.097	0.152
0.75	0.023	0.067	0.118	0.022	0.060	0.139	0.022	0.077	0.130
0.80	0.022	0.069	0.135	0.016	0.062	0.113	0.018	0.068	0.129
0.85	0.037	0.078	0.119	0.016	0.049	0.088	0.019	0.056	0.107
0.90	0.038	0.080	0.118	0.015	0.035	0.073	0.014	0.042	0.079
SW_n	0.040	0.076	0.109	0.011	0.033	0.060	0.015	0.032	0.053
SW_n^{LS}	0.048	0.090	0.127	0.019	0.047	0.077	0.007	0.028	0.044

Table 2: Power - Model 2

| | $n = 250$ | | | $n = 500$ | | | $n = 1000$ | | |
| | | | | Theoretical size | | | | | |
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
0.10	0.473	0.697	0.782	0.918	0.979	0.985	1.000	1.000	1.000
0.15	0.360	0.586	0.695	0.790	0.922	0.961	0.998	1.000	1.000
0.20	0.348	0.555	0.671	0.764	0.902	0.946	0.995	1.000	1.000
0.25	0.331	0.558	0.665	0.752	0.906	0.940	0.991	0.998	1.000
0.30	0.358	0.568	0.672	0.726	0.883	0.941	0.983	0.996	0.999
0.35	0.362	0.559	0.660	0.713	0.852	0.914	0.963	0.991	0.998
0.40	0.348	0.535	0.663	0.695	0.846	0.900	0.955	0.981	0.990
0.45	0.356	0.546	0.666	0.665	0.818	0.885	0.934	0.980	0.989
0.50	0.365	0.550	0.655	0.637	0.813	0.874	0.925	0.969	0.986
0.55	0.346	0.548	0.640	0.637	0.794	0.871	0.920	0.967	0.981
0.60	0.341	0.522	0.623	0.634	0.803	0.873	0.917	0.972	0.986
0.65	0.319	0.507	0.614	0.633	0.802	0.876	0.922	0.975	0.988
0.70	0.297	0.494	0.606	0.633	0.826	0.877	0.931	0.976	0.994
0.75	0.303	0.498	0.613	0.679	0.843	0.896	0.958	0.990	0.995
0.80	0.281	0.483	0.619	0.730	0.868	0.917	0.969	0.992	0.997
0.85	0.283	0.490	0.617	0.759	0.897	0.935	0.983	0.995	0.999
0.90	0.277	0.497	0.619	0.795	0.916	0.948	0.990	0.998	1.000
SW_n	0.284	0.517	0.660	0.856	0.952	0.971	1.000	1.000	1.000
SW_n^{LS}	0.704	0.831	0.889	0.968	0.987	0.992	1.000	1.000	1.000

Table 3: Power - Model 3

| | $n = 250$ | | | $n = 500$ | | | $n = 1000$ | | |
| | Theoretical size | | | | | | | | |
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
0.10	0.978	0.996	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.15	0.940	0.986	0.994	0.999	1.000	1.000	1.000	1.000	1.000
0.20	0.836	0.929	0.964	0.997	1.000	1.000	1.000	1.000	1.000
0.25	0.691	0.823	0.883	0.970	0.989	0.995	1.000	1.000	1.000
0.30	0.540	0.702	0.768	0.876	0.946	0.968	0.991	0.999	0.999
0.35	0.428	0.574	0.676	0.672	0.818	0.886	0.923	0.974	0.983
0.40	0.326	0.470	0.568	0.467	0.638	0.725	0.683	0.835	0.895
0.45	0.245	0.383	0.483	0.309	0.458	0.553	0.381	0.564	0.662
0.50	0.201	0.325	0.397	0.219	0.344	0.426	0.196	0.316	0.408
0.55	0.182	0.294	0.380	0.195	0.314	0.389	0.139	0.238	0.305
0.60	0.179	0.304	0.390	0.206	0.340	0.419	0.193	0.297	0.398
0.65	0.218	0.354	0.452	0.288	0.456	0.552	0.366	0.535	0.629
0.70	0.296	0.435	0.530	0.467	0.638	0.727	0.674	0.824	0.885
0.75	0.392	0.574	0.663	0.685	0.835	0.891	0.910	0.967	0.985
0.80	0.543	0.711	0.804	0.870	0.949	0.978	0.992	0.999	1.000
0.85	0.695	0.840	0.899	0.974	0.994	0.999	1.000	1.000	1.000
0.90	0.828	0.929	0.962	0.997	1.000	1.000	1.000	1.000	1.000
SW_n	0.948	0.985	0.993	1.000	1.000	1.000	1.000	1.000	1.000
SW_n^{LS}	0.063	0.104	0.142	0.029	0.055	0.078	0.011	0.030	0.043

Table 4: Power - Model 4

	n = 250			n = 500			n = 1000		
	Theoretical size								
	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
0.10	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.15	0.132	0.232	0.316	0.244	0.435	0.549	0.657	0.838	0.902
0.20	0.110	0.197	0.268	0.141	0.263	0.356	0.300	0.488	0.598
0.25	0.124	0.209	0.269	0.130	0.225	0.296	0.166	0.285	0.359
0.30	0.124	0.216	0.277	0.146	0.233	0.300	0.139	0.228	0.302
0.35	0.138	0.255	0.338	0.169	0.282	0.367	0.174	0.296	0.368
0.40	0.173	0.298	0.392	0.225	0.361	0.462	0.292	0.464	0.565
0.45	0.222	0.365	0.463	0.345	0.505	0.605	0.519	0.710	0.811
0.50	0.288	0.441	0.561	0.482	0.647	0.752	0.797	0.896	0.943
0.55	0.365	0.553	0.678	0.658	0.836	0.903	0.940	0.980	0.992
0.60	0.499	0.695	0.793	0.844	0.934	0.962	0.990	0.999	1.000
0.65	0.655	0.817	0.874	0.946	0.977	0.993	1.000	1.000	1.000
0.70	0.780	0.909	0.947	0.991	0.999	1.000	1.000	1.000	1.000
0.75	0.888	0.951	0.978	0.999	1.000	1.000	1.000	1.000	1.000
0.80	0.951	0.982	0.995	1.000	1.000	1.000	1.000	1.000	1.000
0.85	0.980	0.998	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.90	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SW_n	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SW_n^{LS}	0.636	0.781	0.841	0.940	0.973	0.981	0.997	0.999	0.999

Table 5: Local Power - Model 5. $n = 250$. $\alpha = 0.05$.

c	0	1	2	3	4	5
0.10	0.123	0.120	0.161	0.205	0.334	0.487
0.15	0.076	0.089	0.122	0.184	0.288	0.381
0.20	0.069	0.085	0.125	0.184	0.278	0.379
0.25	0.060	0.095	0.120	0.167	0.272	0.377
0.30	0.073	0.079	0.124	0.191	0.275	0.377
0.35	0.067	0.074	0.123	0.189	0.288	0.399
0.40	0.075	0.083	0.144	0.188	0.276	0.395
0.45	0.072	0.068	0.126	0.187	0.302	0.400
0.50	0.078	0.090	0.123	0.186	0.276	0.379
0.55	0.067	0.084	0.127	0.174	0.273	0.385
0.60	0.061	0.078	0.099	0.170	0.261	0.363
0.65	0.072	0.071	0.097	0.168	0.238	0.348
0.70	0.061	0.065	0.094	0.139	0.220	0.331
0.75	0.063	0.062	0.081	0.116	0.202	0.286
0.80	0.066	0.058	0.080	0.121	0.188	0.290
0.85	0.065	0.070	0.073	0.100	0.182	0.278
0.90	0.068	0.066	0.076	0.102	0.162	0.271
SW_n	0.077	0.082	0.087	0.089	0.162	0.282
SW_n^{LS}	0.108	0.117	0.184	0.293	0.477	0.641