Journal of Econometrics 152 (2009) 165–178

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Unit root quantile autoregression testing using covariates

Antonio F. Galvao Jr.

Department of Economics, University of Wisconsin-Milwaukee, Bolton Hall 868, 3210 N. Maryland Ave., Milwaukee, WI, 53201, USA

ARTICLE INFO

Article history:
Available online 16 January 2009

JEL classification:
C14
C22
F40

Keywords:
Unit root
Quantile
Covariates
Linear time trend

ABSTRACT

This paper extends unit root tests based on quantile regression proposed by Koenker and Xiao [Koenker, R., Xiao, Z., 2004. Unit root quantile autoregression inference, Journal of the American Statistical Association 99, 775–787] to allow stationary covariates and a linear time trend. The limiting distribution of the test is a convex combination of Dickey–Fuller and standard normal distributions, with weight determined by the correlation between the equation error and the regression covariates. A simulation experiment is described, illustrating the finite sample performance of the unit root test for several types of distributions. The test based on quantile autoregression turns out to be especially advantageous when innovations are heavy-tailed. An application to the CPI-based real exchange rates using four different countries suggests that real exchange rates are not constant unit root processes.

1. Introduction

The unit root hypothesis has important implications for determining the effects of random shocks on economic variables. Recently, methods for detecting the presence of a unit root in semiparametric time series models have attracted interest in both theory and applications, since one way to increase power performance is the use of robust estimators, together with the associated inference apparatus. Such tests are designed to have a good power for many different error distributions. Thompson (2004), Koenker and Xiao (2004), Hasan (2001), Hasan and Koenker (1997), Rothenberg and Stock (1997), Herce (1996) and Lucas (1995) discuss robust estimation and testing in the presence of the unit root process.

Koenker and Xiao (2004) propose new tests of the unit root hypothesis based on the quantile autoregression (QAR) approach in an univariate context. Since many empirical applications have notoriously heavy-tailed behavior, it is important to consider estimation and inference procedures which are robust to departures from Gaussian conditions and are applicable to nonstationary time series. Quantile autoregression methods provide a framework for robust inference and allow one to explore a range of conditional quantiles exposing a variety of forms of conditional heterogeneity. Such models can still deliver important insights about dynamics and persistency in economic time series, and thus provide a useful tool in empirical diagnostic time series analysis. Koenker and Xiao (2004) suggest a t-ratio statistic to test the hypothesis of unit root that accounts only for intercept, and without covariates in the estimated equation. However, following Nelson and Plosser (1982), a common motivation for unit root testing is to test the hypothesis that a series is difference stationary against the alternative that it is trend stationary. Such tests are interesting because, under the alternative hypothesis of stationarity, time series exhibit trend reversion characteristics, whereas under the null they do not.

Hansen (1995) proposes a least squares based covariate augmented Dickey–Fuller (CADF) test, and shows that including correlated stationary covariates in the regression equation can lead to a more precise estimate of the autoregressive coefficient and consequently to large power gains. In this context, Elliott and Jansson (2003) and Pesavento (2007) propose generalizations of the CADF test. Therefore, another important extension of Koenker and Xiao (2004) is the inclusion of at least one covariate when testing for unit root.

This paper aims to generalize the quantile autoregression unit root test by introducing stationary covariates and a linear time trend into the quantile autoregression model. We explore estimation and inference in a model where there is one series that potentially has a unit root, and this series potentially covaries with some available stationary variable. The findings suggest that the limiting distribution of the t-ratio statistic based on quantile regression estimation after adding stationary covariates and a linear time trend continues to be a combination of Dickey–Fuller and Normal distributions, with weights determined by the correlation between the equation error and the regression covariates. Monte Carlo experiments show that the test based on covariate quantile autoregression (CQAR) turns out to be especially advantageous when innovations are non-Gaussian heavy-tailed. In particular, the results show that the quantile autoregression

E-mail address: agalvao@uwm.edu.

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.01.007

test proposed in this paper presents power gains relative to the QAR test proposed by Koenker and Xiao (2004) when there is an available stationary covariate and it is included in the estimated model. In addition, in the non-Gaussian heavy-tailed distribution case, the CQAR unit root test presents more power than the CADF test. Finally, we illustrate the test with an application to the CPI-based real exchange rates using four different countries: Canada, Japan, Switzerland and the United Kingdom. The results indicate that real exchange rates are not constant unit root processes.

The paper is organized as follows. In Section 2, we introduce the model and estimation. Section 3 presents the test and its asymptotic behavior, and in Section 4 we conduct a Monte Carlo experiment to study the performance of the estimator in finite sample. In Section 5 we apply the test to the CPI-based real exchange rates. Finally, Section 6 concludes the paper.

2. Quantile autoregression

2.1. The model and assumptions

The univariate series y_t consists of a deterministic and stochastic component

$$y_t = d_t + S_t, \tag{1}$$

for $t = 1, \dots, n$, where the deterministic component can be: $d_t = 0$, $d_t = \mu_1$, or $d_t = \mu_1 + \mu_2 t$. The stochastic component S_t is modeled as

$$a(L)\Delta S_t = \delta S_{t-1} + e_t, \tag{2}$$

where Δ is the usual difference operator, $a(L) = 1 - a_1 L - a_2 L^2 - \cdots - a_p L^p$ is a pth order polynomial in the lag operator, and

$$e_t = b(L)' x_t + u_t, \tag{3}$$

where x_t is a mean zero υ-vector, and $b(L) = b_{q_2} L^{-q_2} + \cdots + b_{q_1} L^{q_1}$ is a lag polynomial allowing for both leads and lags of x_t to enter the equation for e_t. So, we let the innovations (e_t) in the model be serially correlated, and also allow them to be related to other stationary covariates. We wish to test the unit root hypothesis $\delta = 0$ versus the alternative $\delta < 0$.

Assumptions: for some $p > r > 2$

A1. $\{u_t, x_t\}$ is covariance stationary and strong mixing with mixing coefficients α_m, which satisfies $\sum_{m=1}^{\infty} \alpha_m^{1/r-1/p} < \infty$;
A2. $\sup_t E[|x_t|^p + |u_t|^p] < \infty$;
A3. $E[x_{t-k} u_t] = 0$ for $q_1 \le k \le q_2$;
A4. $E[u_t u_{t-k}] = 0$ for $k \ge 1$;
A5. $E(\phi_t \phi_t') > 0$, where $\phi_t = (\Delta y_{t-1}, \dots, \Delta y_{t-q}, x_{t-q_2}, \dots, x_{t+q_1})'$;
A6. the roots of $a(L)$ all lie outside the unit circle;
A7. the distribution function of u_t, F, has differentiable continuous Lebesgue density, $0 < f(u) < \infty$, with bounded derivatives f' on $\{u : 0 < F(u) < 1\}$.

Assumptions A1–A6 are the same as in Hansen (1995). A1 and A2 state weak dependence and moment restrictions. Assumptions A3 and A4 exclude linear dependence. Assumption 6 is a typical stationarity assumption. Finally, assumption A7 is a standard assumption in quantile regression literature and imposes restriction on the density function of u_t.

2.2. Estimation

Estimation and testing are based on the following linear model[1]:

$$y_t = \mu_1 + \mu_2 t + \alpha y_{t-1} + \sum_{j=1}^{p} \alpha_j \Delta y_{t-j} + \sum_{l=-q_1}^{q_2} \gamma_l x_{t-l} + u_t. \tag{4}$$

The model may thus be written as

$$Q_{y_t}(\tau|\Im_{t-1}) = \mu_1 + \mu_2 t + \alpha y_{t-1} + \sum_{j=1}^{p} \alpha_j \Delta y_{t-j} + \sum_{l=-q_1}^{q_2} \gamma_l x_{t-l} + F_u^{-1}(\tau)$$

where Q_{y_t} denotes the τ-th conditional quantile of y_t conditional on $\Im_{t-1}$, where $\Im_{t-1}$ is the σ-field generated by $\{u_s, s < t, x_{t-q_2}, \dots, x_{t+q_1}\}$. The F_u denotes the common distribution function of the errors. Let $\mu_1(\tau) = \mu_1 + F_u^{-1}(\tau)$, and define

$$z_t = (1, t, y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t-p}, x_{t-q_2}, \dots, x_{t+q_1})'$$
and $\beta(\tau) = (\mu_1(\tau), \mu_2, \alpha, \alpha_1, \dots, \alpha_p, \gamma_{q_1}, \dots, \gamma_{-q_2})'$,

thus, we have

$$Q_{y_t}(\tau|\Im_{t-1}) = z_t' \beta(\tau). \tag{5}$$

Estimation of the linear quantile autoregression model involves solving the problem

$$\min_{\beta \in R^{3+p+q}} \sum_{t=1}^{n} \rho_\tau \left(y_t - z_t' \beta \right), \tag{6}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ as in Koenker and Bassett (1978). We shall be concerned with the limiting distribution of the coefficients in (6), more specifically with $\hat{\alpha}(\tau)$ and its t-ratio statistic under the hypothesis of unit root. Thus under the null hypothesis $\alpha(\tau) = 1$.

2.3. QAR asymptotics under unit root hypothesis

In this section we describe the limiting distribution of the quantile autoregression process under the unit root hypothesis, where the observations $\{y_t\}_{t=1}^n$ come from the data generating process as (1)–(3) with $d_t = \mu_1 + \mu_2 t$.[2] First, in order to derive the asymptotic properties of $\hat{\alpha}$, and without loss of generality, we use a convenient reparametrization of the objective function, by applying the quantile equivariance property. Further, we derive the asymptotic distribution of $\hat{\alpha}$.

Consider the estimator $\hat{\alpha}$ which solves

$$\min_{\beta \in R^{3+p+q}} \sum_{t=1}^{n} \rho_\tau \left(y_t - \mu_1 - \mu_2 t - \alpha y_{t-1} - \sum_{j=1}^{p} \alpha_j \Delta y_{t-j} - \sum_{l=-q_1}^{q_2} \gamma_l x_{t-l} \right). \tag{7}$$

Define $\tilde{y}_t = y_t - \mu_1 - \mu_2 t$. According to the equivariance property, Theorem 3.2 part 4 in Koenker and Bassett (1978), $\tilde{\beta}(\tau, y, XA) = A^{-1} \tilde{\beta}(\tau, y, X)$, hence solving (7) is equivalent to

$$\min_{\beta \in R^{3+p+q}} \sum_{t=1}^{n} \rho_\tau \left(y_t - \eta - \theta t - \alpha \tilde{y}_{t-1} - \sum_{j=1}^{p} \alpha_j \Delta y_{t-j} - \sum_{l=-q_1}^{q_2} \gamma_l x_{t-l} \right), \tag{8}$$

where $\eta = \mu_1 + \alpha(\mu_2 - \mu_1)$, $\theta = \mu_2 + \alpha\mu_2$, and $\tilde{y}_{t-1} = y_{t-1} - \mu_1 - \mu_2(t-1)$. Therefore, $\hat{\alpha}$ which solves the minimization problem (7) also solves the minimization problem (8) and we can describe the asymptotic properties of $\hat{\alpha}$ based on the latter equation. Thus, let

$$\tilde{z}_t = (1, t, \tilde{y}_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t-p}, x_{t-q_2}, \dots, x_{t+q_1})' \tag{9}$$
and $\beta(\tau) = (\eta, \theta, \alpha, \alpha_1, \dots, \alpha_p, \gamma_{q_1}, \dots, \gamma_{-q_2})'$.

[1] When $d_t = \mu_1 + \mu_2 t$ the model (1)–(3) can be written as $a(L)\Delta y_t = \mu_1^* + \mu_2^* t + \delta y_{t-1} + b(L) x_t + u_t$, where $\mu_1^* = a(1)\mu_2 - \delta\mu_1$ and $\mu_2^* = -\delta\mu_2$. Since we have interest only in the autoregression coefficient, we omit the superscript.

[2] All results may be extended to the model generated by $d_t = \mu_1$ or $d_t = 0$.

The components in the vector $\hat{\beta}(\tau)$ have different rates of convergence. Therefore, it is useful to define the diagonal matrix $D_n = diag\left(n^{\frac{1}{2}}, n^{\frac{3}{2}}, n, n^{\frac{1}{2}}, n^{\frac{1}{2}}, \ldots, n^{\frac{1}{2}}\right)$ for standardization purposes. Denote $\hat{v} = D_n(\hat{\beta}(\tau) - \beta(\tau))$, and write $\rho_\tau(y_t - \hat{\beta}(\tau)'\tilde{z}_t)$ as $\rho_\tau(u_{t\tau} - (D_n^{-1}\hat{v})'\tilde{z}_t)$ where $u_{t\tau} = y_t - \tilde{z}_t'\beta(\tau)$. Minimization of (8) is equivalent to the following problem:

$$\min_v \sum_{t=1}^n [\rho_\tau(u_{t\tau} - (D_n^{-1}v)'\tilde{z}_t) - \rho_\tau(u_{t\tau})]. \tag{10}$$

If $\hat{v}$ is a minimizer of $H_n(v) = \sum_{t=1}^n [\rho_\tau(u_{t\tau} - (D_n^{-1}v)'\tilde{z}_t) - \rho_\tau(u_{t\tau})]$, we have $\hat{v} = D_n(\hat{\beta}(\tau) - \beta(\tau))$. The objective function $H_n(v)$ is a convex random function. Knight (1989, 1991) and Pollard (1991) show that if the finite-dimensional distributions of $H_n(\cdot)$ converges weakly to those of $H(\cdot)$, and if $H(\cdot)$ has a unique minimum, the convexity of $H_n(\cdot)$ implies that $\hat{v}$ converges in distribution to the minimizer of $H(\cdot)$. Denoting $\psi_\tau(u) = \tau - I(u < 0)$ for $u \neq 0$ and following the approach of Knight (1989) the objective function for the minimization problem (10) can be rewritten as

$$H_n(v) = -\sum_{t=1}^n v'D_n^{-1}\tilde{z}_t\psi_\tau(u_{t\tau}) + \sum_{t=1}^n \int_0^{(D_n^{-1}v)'\tilde{z}_t} \{I(u_{t\tau} \le s) - I(u_{t\tau} < 0)\}\mathrm{d}s.$$

Therefore, in order to derive the asymptotic results for the limiting distribution of $D_n(\hat{\beta}(\tau) - \beta(\tau))$ we need to study the convergence of the two terms of $H_n(v)$ in the above equation. Thus, once we show that $H_n(\cdot)$ converges weakly to $H(\cdot)$ we just need to find the minimizer of $H(\cdot)$, and $\hat{v}$ converges in distribution to that minimizer.

By definition of $u_{t\tau}$, we have that $E[\psi_\tau(u_{t\tau})|\Im_{t-1}] = 0$. Both $u_{t\tau}$ and $\psi_\tau(u_{t\tau})$ have mean zero, and are correlated. Then, under the unit root hypothesis and Assumptions A1–A6, the partial sums of the vector process $(e_t, \psi_\tau(u_{t\tau}))$ follow a bivariate invariance principle (Hansen, 1992b):

$$n^{-1/2}\sum_{t=1}^{[nr]}(e_t, \psi_\tau(u_{t\tau}))' \Rightarrow \left(B_e(r), B_\psi^\tau(r)\right) = BM(0, \Sigma(\tau))$$

where

$$\Sigma(\tau) = \begin{bmatrix} \sigma_e^2(\tau) & \sigma_{e\psi}(\tau) \\ \sigma_{e\psi}(\tau) & \sigma_\psi^2(\tau) \end{bmatrix} \tag{11}$$

is the long run covariance matrix and can be written as $\Sigma_0(\tau) + \Sigma_1(\tau) + \Sigma_1'(\tau)$, where $\Sigma_0(\tau) = E\left[(e_t, \psi_\tau(u_{t\tau}))'(e_t, \psi_\tau(u_{t\tau}))\right]$ and $\Sigma_1(\tau) = \sum_{s=2}^\infty E\left[(e_1, \psi_\tau(u_{1\tau}))'(e_s, \psi_\tau(u_{s\tau}))\right]$. We define $\delta^2 = \frac{\sigma_{e\psi}^2(\tau)}{\sigma_e^2(\tau)\sigma_\psi^2(\tau)}$ and assume that $\delta^2 < 1$.[3]

The random function $n^{-1/2}\sum_{t=1}^{[nr]}\psi_\tau(u_{t\tau})$ converges weakly to a two parameter process $B_\psi^\tau(r) = B_\psi(\tau, r)$.[4] The result follows from finite dimensional convergence and tightness. The first result has been long established in the literature, and can be found, for instance, in Koenker (2005). We show tightness in Lemma A2 in the Appendix. Thus, we obtain that the limiting variate $B_\psi^\tau(r)$, viewed as a random function of τ, is a Brownian bridge over $\tau \in \mathcal{T}$, where $\mathcal{T} \equiv [\tau_0, 1-\tau_0]$, $\tau_0 \in (0, 1/2)$. Hence, the two parameter process $B_\psi^\tau(r)$ is partially Brownian motion and partially Brownian bridge in the sense that for fixed r, $B_\psi^\tau(r) = B_\psi(\tau, r)$ is a rescaled Brownian bridge, while for each τ, $n^{-1/2}\sum_{t=1}^{[nr]}\psi_\tau(u_{t\tau})$ converges weakly to a Brownian motion with variance $\tau(1-\tau)$. Thus, for a fixed pair (τ, r), $B_\psi^\tau(r) = B_\psi(\tau, r) \sim N(0, \tau(1-\tau)r)$. In addition, as argued by Park and Phillips (1988) $n^{-3/2}\sum_{t=1}^n r\psi_\tau(u_{t\tau}) \Rightarrow \int r\mathrm{d}B_\psi^\tau$.

The next three lemmas provide asymptotic results that are useful in deriving the limiting distribution of $D_n(\hat{\beta}(\tau) - \beta(\tau))$. The results are derived as processes indexed by $\tau \in \mathcal{T}$.

Lemma 1. *Under Assumptions* A1–A6*,*

$$n^{-1}\sum_{t=1}^n \tilde{y}_{t-1}\psi_\tau(u_{t\tau}) \Rightarrow a(1)^{-1}\int_0^1 B_e\mathrm{d}B_\psi^\tau.$$

Lemma 2. *Under Assumptions* A1–A6*,*

$$D_n^{-1}\sum_{t=1}^n \tilde{z}_t\psi_\tau(u_{t\tau}) \Rightarrow \begin{bmatrix} \int_0^1 \bar{B}_e\mathrm{d}B_\psi^\tau \\ \Phi_\tau \end{bmatrix} := \Phi_\tau^*$$

where $\bar{B}_e(r) = \left[1, r, a(1)^{-1}B_e(r)\right]'$, and $\Phi_\tau = [\Phi_1, \ldots, \Phi_{p+q}]$ is a p+q-dimensional normal variate with covariance matrix $\tau(1-\tau)\Omega_\Phi$, where the elements of Ω_Φ are the elements of the matrix $E[\phi_t\phi_t']$, and Φ_τ is independent with $\int_0^1 \bar{B}_e\mathrm{d}B_\psi^\tau$.

Lemma 3. *Under Assumptions* A1–A7*,*

$$\sum_{t=1}^n \int_0^{(D_n^{-1}v)'\tilde{z}_t} \{I(u_{t\tau} \le s) - I(u_{t\tau} < 0)\}\mathrm{d}s \Rightarrow \frac{1}{2}f(F^{-1}(\tau))v'\left[\int_0^1 \bar{B}_e\bar{B}_e'\right]v.$$

The limiting distribution of the quantile autoregression estimator for the unit root model is given by Proposition 1. The proof of the proposition is given in the Appendix as well as the proof of the three lemmas stated previously.

Proposition 1. *Let y_t be determined by (1)–(3). Under unit root hypothesis and Assumptions* A1–A7*,*

$$D_n(\hat{\beta}(\tau) - \beta(\tau)) \Rightarrow \frac{1}{f(F^{-1}(\tau))}\begin{bmatrix} \int_0^1 \bar{B}_e\bar{B}_e' & 0_{3\times(p+q)} \\ 0_{(p+q)\times 3} & \Omega_\Phi \end{bmatrix}^{-1} \times \begin{bmatrix} \int_0^1 \bar{B}_e\mathrm{d}B_\psi^\tau \\ \Phi_\tau \end{bmatrix}. \tag{12}$$

Since $\left[\int_0^1 \bar{B}_e\bar{B}_e'\right]^{-1}$ is a non-singular matrix, and we are interested in the asymptotic behavior of $\hat{\alpha}(\tau)$, we define $e_z = (0, 0, 1, 0, \ldots, 0)$. Therefore, we have $n(\hat{\alpha}(\tau) - 1) = e_zD_n(\hat{\beta}(\tau) - \beta(\tau))$. In addition, it is more convenient to use the techniques of Park and Phillips (1988) and formulate the above convergence in terms of detrended Brownian motions. As a consequence of the above proposition, we have the following corollary, formulated in terms of detrended Brownian motions, which is useful for the derivation of the asymptotic distribution of the test for the unit root hypothesis. The proof is given in the Appendix.

[3] As in Elliott and Jansson (2003), δ^2 is assumed to be strictly less than one, thus ruling out the possibility that under the null, the partial sum of y_t cointegrates with x_t.

[4] Similar results for weak convergence to a two parameter process in the quantile regression literature can be found in Su and Xiao (in press) and Qu (2008).

Corollary 1. *Under the assumptions of Proposition 1,*

$$n(\hat{\alpha}(\tau)-1) \Rightarrow \frac{a(1)}{f(F^{-1}(\tau))}\left[\int_0^1 \underline{B}_e^2\right]^{-1}\left[\int_0^1 \underline{B}_e \mathrm{d}B_\psi^\tau\right]$$

where $\underline{B}_e(r) = B_e(r) - \left[\int_0^1 (4-6s)B_e - r\int_0^1 (6-12s)B_e\right]$ *is the detrended version of the Brownian motion* B_e.

3. Inference on quantile autoregression

Inference based on the quantile autoregression provides a more robust approach to testing the unit root hypothesis. Including related stationary covariates in the model leads to gains in power. In addition, considering a deterministic time trend in the regression is useful because under the alternative hypothesis of stationarity, time series exhibit trend reversion characteristics. We consider the same t-ratio test as Koenker and Xiao (2004),

$$t_n(\tau) = \frac{\widehat{f(F^{-1}(\tau))}}{\sqrt{\tau(1-\tau)}}\left(Y_{-1}' M_Z Y_{-1}\right)^{1/2}(\hat{\alpha}(\tau)-1), \tag{13}$$

where $\widehat{f(F^{-1}(\tau))}$ is a consistent estimator of $f(F^{-1}(\tau))$, Y_{-1} is a vector of lagged dependent variables (y_{t-1}) and M_Z is the projection matrix onto the space orthogonal to $Z = (1, t, \Delta y_{t-1}, \ldots, \Delta y_{t-p}, x_{t-q_2}, \ldots, x_{t+q_1})$. The following proposition shows the convergence of t-ratio $t_n(\tau)$, $\tau \in \mathcal{T} \equiv [\tau_0, 1-\tau_0]$, $\tau_0 \in (0, 1/2)$, proposed in Eq. (13). The proof is given in the Appendix.

Proposition 2. *Under the unit root hypothesis, Assumptions* A1–A7, *and using the results of the previous section,*

$$t_n(\tau) \Rightarrow t(\tau) = \frac{1}{\sqrt{\tau(1-\tau)}}\left[\int_0^1 \underline{B}_e^2\right]^{-1/2}\left[\int_0^1 \underline{B}_e \mathrm{d}B_\psi^\tau\right]. \tag{14}$$

At any fixed τ, the test statistic $t_n(\tau)$ is simply the quantile regression counterpart of the well-know Dickey–Fuller t-ratio test for unit root. Therefore, unit root tests may be constructed based on some selected quantiles, such as median, lower quantiles, upper quantiles, etc.

3.1. Decomposing the limiting distribution of $t_n(\tau)$

The limiting distribution of $t_n(\tau)$ is nonstandard and depends on nuisance parameters $(\sigma_e^2(\tau), \sigma_{\psi,e}(\tau))$ since B_e and B_ψ^τ are correlated Brownian motions. But the limiting distribution of $t_n(\tau)$ can be decomposed in the linear combination of two independent contributions, with weights determined by a long-run correlation coefficient that can be consistently estimated.

Following Phillips and Hansen (1990) and Phillips (1995) we have the following decomposition for the numerator of Eq. (14),

$$\int \underline{B}_e \mathrm{d}B_\psi^\tau = \int \underline{B}_e \mathrm{d}B_{\psi,e}^\tau + \lambda_{e\psi}(\tau)\int \underline{B}_e \mathrm{d}B_e$$

where $\lambda_{e\psi}(\tau) = \sigma_{e\psi}(\tau)/\sigma_e^2(\tau)$, and $B_{\psi,e}^\tau$ is a Brownian motion with variance $\sigma_{\psi,e}^2(\tau) = \sigma_\psi^2(\tau) - \sigma_{e\psi}^2(\tau)/\sigma_e^2(\tau)$ and independent of $\underline{B}_e$.[5] Therefore, the limiting distribution of $t_n(\tau)$ can be rewritten as

$$t(\tau) = \frac{1}{\sqrt{\tau(1-\tau)}}\frac{\int \underline{B}_e \mathrm{d}B_{\psi,e}^\tau}{\left(\int \underline{B}_e^2\right)^{1/2}} + \frac{\lambda_{e\psi}(\tau)}{\sqrt{\tau(1-\tau)}}\frac{\int \underline{B}_e \mathrm{d}B_e}{\left(\int \underline{B}_e^2\right)^{1/2}}.$$

Noting that $\sigma_\psi^2(\tau) = \tau(1-\tau)$ and the fact that W_2 is independent of both W_1 and $\underline{W}_1$ implies that $\int \underline{W}_1 \mathrm{d}W_2 / \left(\int (\underline{W}_1)^2\right)^{1/2}$ is distributed as $N(0,1)$, so one can show that the limiting distribution of $t_n(\tau)$ can be written as,

$$t(\tau) = \delta\left(\int_0^1 \underline{W}_1^2\right)^{-1/2}\int_0^1 \underline{W}_1 \mathrm{d}W_1 + \sqrt{1-\delta^2}N(0,1), \tag{15}$$

where

$$\delta = \delta(\tau) = \frac{\sigma_{e\psi}(\tau)}{\sigma_e \sigma_\psi(\tau)} = \frac{\sigma_{e\psi}(\tau)}{\sigma_e\sqrt{\tau(1-\tau)}}.$$

Therefore, the limiting distribution of the $t_n(\tau)$ is a mixture of a Dickey–Fuller component, $\left(\int_0^1 \underline{W}_1^2\right)^{-1/2}\int_0^1 \underline{W}_1 \mathrm{d}W_1$, and a standard normal component, with the weights determined by the parameter δ.[6]

The limiting distribution (15) is the same as that of the CADF test of Hansen (1995) replacing ρ by δ. We reproduce the tables of critical values for demeaned and detrended cases in the Appendix as Table A. Note that we find the same asymptotic distribution for $t_n(\tau)$ as in Koenker and Xiao (2004). The form of the limiting distribution of the t-ratio statistic based on the quantile regression estimation after adding covariates and a linear time trend, which was already a mix of DF and normal distribution, does not change. However, the weight δ does, obviously, carry the information contained in the new covariates. In addition, it is important to note that the test proposed in this paper is a generalization of the QAR unit root test and, in fact, it has the same asymptotic distribution of QAR unit root test when there is no information in the stationary covariates, i.e. when the correlation between the stationary covariate and the variable being tested is zero. In order to see this, note that when $b(L)' = 0$ in Eq. (3), $e_t = u_t$. Therefore, $\delta(\tau) = \frac{\sigma_{e\psi}(\tau)}{\sigma_e\sigma_\psi(\tau)} = \frac{\sigma_{u\psi}(\tau)}{\sigma_u\sigma_\psi(\tau)}$ as in Koenker and Xiao (2004) and we have the same limiting distribution for $t_n(\tau)$.

In order to calculate the test statistic and to select the appropriate critical values from the table, it is necessary to obtain consistent estimates of various nuisance parameters. There is a large amount of literature on estimating $f(F^{-1}(\tau))$. Noticing that $\mathrm{d}F^{-1}(t)/\mathrm{d}t = \left(f(F^{-1}(t))\right)^{-1}$, it is natural to use the estimator

$$f_n(F_n^{-1}(\tau)) = \frac{2h_n}{F_n^{-1}(t+h_n) - F_n^{-1}(t-h_n)},$$

where $F_n^{-1}(s)$ is an estimate of $F^{-1}(s)$ and h_n is a bandwidth which tends to zero as $n \to \infty$. Bassett and Koenker (1982) show that $\hat{Q}(\tau|\bar{z}) = \bar{z}'\hat{\beta}(\tau)$ is a consistent estimator for $F^{-1}(s)$. Therefore, following Koenker and Xiao (2004), the density $f(F^{-1}(\tau))$ can be consistently estimated by

$$f_n(F_n^{-1}(\tau)) = \frac{2h_n}{\bar{z}'(\hat{\beta}(\tau+h_n) - \hat{\beta}(\tau-h_n))}$$

where h_n is a bandwidth which tends to zero as $n \to \infty$, $\hat{\beta}(\tau)$ are the estimates of model (4), and $\bar{z}$ is a vector of averages of z_t. As in Hansen (1995), the long-run variance and covariance parameters can be consistently estimated by kernel estimators

[5] Note that, we can write the Brownian motions $B_e(r)$, $\underline{B}_e$, and $B_{\psi,e}^\tau(r)$ as: $B_e(r) = \sigma_e W_1(r)$; $B_{\psi,e}^\tau(r) = \sigma_{\psi,e}W_2(r)$; $\underline{B}_e(r) = \sigma_e \underline{W}_1(r)$; $\underline{W}_1(r) = W_1(r) - \left[\int_0^1 (4-6s)W_1 - r\int_0^1 (6-12s)W_1\right]$, where $W_1(r)$ and $W_2(r)$ are standard Brownian motions and are independent of one another.

[6] It is possible to relax Assumption A7 of identically distributed innovations. However, the limiting distribution will then depend on the density function of the innovations as a nuisance parameter and will no longer be the linear combination of Dickey–Fuller and normal distributions. As a result, implementation of the test would require extensive simulation of the limiting distribution.

$$\hat{\sigma}_e^2 = \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) C_{ee}(h), \qquad \hat{\sigma}_{e\psi} = \sum_{h=-M}^{M} k\left(\frac{h}{M}\right) C_{e\psi}(h)$$

where, for all $x \in \mathbb{R}$, $|k(x)| \leq 1$ and $k(x) = k(-x)$; $k(0) = 1$; $k(x)$ is continuous at zero and for almost all $x \in \mathbb{R}$; $\int_{\mathbb{R}} |k(x)|\mathrm{d}x \leq \infty$, M is the bandwidth (truncation) parameter satisfying the property that $M \to \infty$ and for some $q \in (1/2, \infty)$, $M^{1+2q}/n = O(1)$ as the sample size $n \to \infty$. The quantities $C_{ee}(h)$, and $C_{e\psi}(h)$ are sample covariances defined by $C_{ee}(h) = n^{-1}\sum' \hat{e}_t\hat{e}_{t+h}$, and $C_{e\psi}(h) = n^{-1}\sum' \hat{e}_t\psi(\hat{u}_{(t+h)\tau})$, where $\sum'$ means summation over $1 \leq t, t+h \leq n$, and $(\hat{e}_t, \psi(\hat{u}_{(t+h)\tau}))$ are the estimates of $(e_t, \psi(u_{(t+h)\tau}))$ from the appropriate regression model.[7] Therefore, consistency of $\hat{\sigma}_e^2$ and $\hat{\sigma}_{e\psi}$ follows from Hansen (1992a), and consistency of δ follows from continuous mapping theorem.

3.2. Kolmogorov–Smirnov test

Another approach to test the unit root hypothesis is to examine the unit root property over a range of quantiles $\tau \in \mathcal{T}$, instead of focusing only on a selected quantile. Following Koenker and Xiao (2004), we may construct a Kolmogorov–Smirnov (KS) type test on the regression quantile process for $\tau \in \mathcal{T} = [\tau_0, 1-\tau_0]$ for some $0 < \tau_0 < \frac{1}{2}$. It is important to note that assumption A7 effectively imposes identically distributed innovations, implying that the KS test can be viewed as a diagnostic tool for the adequacy of the proposed model.[8] We suggest the following quantile regression based statistic for testing the null hypothesis of a unit root:

$$QKS_n = \sup_{\tau \in \mathcal{T}} |t_n(\tau)|.$$

In practice, one may calculate $t_n(\tau)$ at, say $\{\tau_i = i/n\}_{i=1}^n$, and then the statistic QKS_t can be constructed by taking the maximum over $\tau_i \in \mathcal{T}$. The limiting distribution of the test is given by the following proposition

Proposition 3. *Under Assumptions* A1–A7 *and unit root hypothesis*

$$\sup_{\tau \in \mathcal{T}} |t_n(\tau)| \Rightarrow \sup_{\tau \in \mathcal{T}} |t(\tau)|.$$

Proof. This Proposition follows from Proposition 2 and the continuous mapping theorem. ■

We develop inferential methods to obtain the critical values for the QKS_n by using simulation methods. The asymptotic null distribution of the CQAR t-ratio test depends on nuisance parameters, but for fixed τ, the critical values can be calculated and tabulated by holding the parameter $\delta(\tau)$ fixed and simulating the components of the limiting distribution, as in Hansen (1995). Seo (1999) simulates draws of $t(\tau)$ for each δ and use the empirical quantiles as critical values, and Shin and So (1999) also describe simulation-based procedures.

For the Kolmogorov–Smirnov test, the critical values cannot be tabulated so simply. The asymptotically correct critical value, $q_t(\delta, \lambda)$, is the $(1-\lambda)$-th quantile of $\sup_{\tau \in \mathcal{T}} |t(\tau)|$, for a prescribed size level λ. For a given test, $\delta(\tau)$ must be estimated so that the critical value of the asymptotic null distribution can be simulated.

We suggest a simulation strategy to compute the critical values for the KS test under the null hypothesis of unit root, where the asymptotic critical values are approximated by the quantile of the empirical distribution, which is based on simulations. The simulation approach is based on the availability of a consistent estimator of $\delta(\tau)$. The procedure can be implemented using the following steps:

(1) For each realization of the Monte Carlo experiment, first compute the estimates of the nuisance parameter $\delta(\tau)$ for $\{\tau_i = i/n\}_{i=1}^n$;

(2) Now, for each $\delta(\tau_i)$, simulate one realization from the Dickey–Fuller and standard normal distributions independently, and compute $t(\tau_i) = \delta(\tau_i)\left(\int_0^1 \underline{W}_1^2\right)^{-1/2} \int_0^1 \underline{W}_1 \mathrm{d}W_1 + \sqrt{1-\delta(\tau_i)^2}$ $N(0, 1)$, and finally, take the maximum of the absolute values over τ_i;

(3) Repeating the last step many times, one can compute the critical value as the corresponding quantile of interest from the empirical distribution of the suprema.

We reject the null hypothesis if the test statistic is greater than the calculated critical value, that is, $QKS_t > q_t(\delta, \lambda)$. An advantage of this method for generating critical values is that it avoids solving the linear programming in each repetition. In the next section we present a Monte Carlo experiment showing evidence of the probability of making type I error and the empirical power, using critical values from this simulation procedure.

4. Monte Carlo

In this section, we investigate the finite sample performance of the presented unit root test by means of the Monte Carlo simulation. The simulation considers the following data generating process (DGP):

$$y_t = \mu_1 + \mu_2 t + \alpha y_{t-1} + e_t$$

where the vector $\xi_t = (e_t, x_t)$ is generated from a VAR model

$$\begin{pmatrix} e_t \\ x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} e_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

The innovations $(\varepsilon_{1t}, \varepsilon_{2t})'$ are *i.i.d.* with covariance matrix,

$$\Sigma = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{21} & 1 \end{pmatrix}.$$

To study the sensitivity of the results, two different effective sample sizes are considered, say $n = 100$ and $n = 200$ (the first 100 observations are discarded to eliminate the start-up effects). Since the results are qualitatively similar, we present only the results for $n = 100$. Three different distributions for innovations $(\varepsilon_{1t}, \varepsilon_{2t})'$ are considered, say standard normal ($N(0, 1)$), student-t distribution with 2 degrees of freedom (t_2), and student-t distribution with 3 degrees of freedom (t_3). The values for α can be: 1.0, 0.95, and 0.9. Note that, when $\alpha = 1$, the rejection rate gives the probability of making a Type I error, and other cases give the empirical power. For each test, the number of repetitions is 2000.

As in Hansen (1995) throughout the experiment we set $a_{22} = 0$ and $a_{11} = 0$, because these parameters do not affect the nuisance parameter δ^2. Therefore, there are three free parameters to set, $(\sigma_{21}, a_{12}, a_{21})$. These parameters control the degree of correlation between e_t and x_t. The first experiment set all three equal to 0, and the remaining 16 set $\sigma_{21} = 0.4$ and varied a_{12} and a_{21} among $\{-0.3, 0, 0.3, 0.6\}$. Since the model is invariant to the presence of a linear time trend, we also set $\mu_1 = 0$ and $\mu_2 = 0$. In this setup we estimate the following linear model with covariates and a linear time trend

$$y_t = \mu_1 + \mu_2 t + \alpha y_{t-1} + \alpha_1 \Delta y_{t-1} + \gamma_1 x_t + \gamma_2 x_{t-1} + u_t.$$

7 For example, under model (4),

$$\hat{u}_t = y_t - \hat{\mu}_1 - \hat{\mu}_2 t - \hat{\alpha} y_{t-1} - \sum_{j=1}^{p} \hat{\alpha}_j \Delta y_{t-j} - \sum_{l=-q_1}^{q_2} \hat{\gamma}_l x_{t-l}$$

$$\text{and} \quad \hat{e}_t = \sum_{l=-q_1}^{q_2} \hat{\gamma}_l x_{t-l} + \hat{u}.$$

8 We thank an anonymous referee for bringing this point to our attention.

Table 1
Probability of error type I for VAR innovations.

a_{12}	a_{21}	$N(0, 1)$		t_2		t_3	
		CADF	CQAR	CADF	CQAR	CADF	CQAR
0.000	0.000	0.052	0.048	0.043	0.046	0.042	0.047
−0.300	−0.300	0.051	0.049	0.048	0.047	0.063	0.064
	0.000	0.051	0.045	0.049	0.036	0.045	0.048
	0.300	0.051	0.046	0.054	0.052	0.057	0.042
	0.600	0.047	0.038	0.051	0.041	0.061	0.037
0.000	−0.300	0.041	0.046	0.044	0.057	0.049	0.047
	0.000	0.042	0.038	0.041	0.037	0.037	0.039
	0.300	0.046	0.031	0.038	0.041	0.050	0.041
	0.600	0.040	0.037	0.049	0.043	0.028	0.032
0.300	−0.300	0.044	0.047	0.047	0.033	0.035	0.047
	0.000	0.047	0.044	0.041	0.039	0.046	0.043
	0.300	0.047	0.032	0.040	0.035	0.056	0.045
	0.600	0.045	0.043	0.042	0.038	0.060	0.044
0.600	−0.300	0.037	0.034	0.041	0.036	0.037	0.040
	0.000	0.039	0.040	0.039	0.042	0.040	0.034
	0.300	0.044	0.033	0.055	0.036	0.049	0.030
	0.600	0.065	0.049	0.051	0.034	0.044	0.035

Table 2
Finite sample power for VAR innovations ($\alpha = 0.95$).

a_{12}	a_{21}	$N(0, 1)$		t_2		t_3	
		CADF	CQAR	CADF	CQAR	CADF	CQAR
0.000	0.000	0.143	0.108	0.122	0.325	0.109	0.144
−0.300	−0.300	0.108	0.102	0.118	0.348	0.099	0.174
	0.000	0.146	0.098	0.131	0.331	0.112	0.171
	0.300	0.131	0.094	0.162	0.336	0.137	0.180
	0.600	0.148	0.135	0.169	0.332	0.133	0.165
0.000	−0.300	0.103	0.097	0.133	0.320	0.113	0.175
	0.000	0.121	0.096	0.175	0.368	0.124	0.176
	0.300	0.185	0.133	0.200	0.437	0.159	0.209
	0.600	0.203	0.123	0.256	0.477	0.200	0.254
0.300	−0.300	0.130	0.114	0.137	0.345	0.113	0.184
	0.000	0.203	0.158	0.237	0.455	0.218	0.272
	0.300	0.315	0.185	0.351	0.566	0.307	0.372
	0.600	0.467	0.236	0.484	0.692	0.469	0.491
0.600	−0.300	0.150	0.133	0.186	0.388	0.133	0.214
	0.000	0.305	0.225	0.338	0.587	0.305	0.384
	0.300	0.527	0.316	0.534	0.767	0.515	0.569
	0.600	0.783	0.487	0.762	0.896	0.747	0.789

We report the results for the t-ratio $t_n(\tau)$ test based in quantile regression at $\tau = 0.50$, using critical values in Table A for the detrended case at 5% level. For comparison reasons, we also report results of the CADF test. The bandwidth parameter, required for the estimation of $f(F^{-1}(\tau))$, was estimated using the Bofinger (1975) bandwidth. For estimation of long-run variance and covariance parameters (σ_e^2, and $\sigma_{e\psi}$) we used Bartlett and Quadratic Spectral windows in the kernel estimators. Since the results are very similar, we present only the results for the Quadratic Spectral kernel estimator.

Table 1 reports the probability of rejecting the null hypothesis under unit root case for the Gaussian, t_2 and t_3 innovations. The results show that the test based on covariate quantile autoregression (CQAR) presents less rejection in finite sample size than the CADF test, especially when the covariates are highly correlated. More precisely, when the parameter a_{12} are $(-0.3, 0.6)$ and the parameter a_{21} are $(0.6, -0.3)$. The downward distortions of the CADF test are mainly for $(a_{12}, a_{21}) = (0.6, -0.3)$ and $(a_{12}, a_{21}) = (0.6, 0)$.

Table 2 reports the probability of rejecting the null hypothesis for the case where $\alpha = 0.95$ with Gaussian, t_2 and t_3 innovations. Results in the table indicate that the CQAR based procedures are superior in the presence of heavy-tailed disturbances. While the CADF power does not change significantly with different distributions, for the CQAR test there is a significant improvement in the finite sample powers when innovations are t_2 and t_3. Indeed, the results in Table 2 show a drastic gain in power of the CQAR test over the CADF test. The gains are very significant, even for the cases where both parameters a_{12} and a_{21} are small. The results for $\alpha = 0.9$ are qualitatively similar and we omit them to save space.

In Table 3 we compare the empirical power for the QAR and the CQAR models when $\alpha = 0.95$ and with Normal errors. The third column shows the results for the model estimated without any covariates (QAR), and in the fourth column the model was estimated including covariates (CQAR) in the regression equation. We can see from Table 3 that, if the DGP is as given above, there is a large gain in power when using the covariates in the estimated model, especially when the variables are very correlated.

We also compute the probability of rejecting the null hypothesis under unit root and the empirical power for the Kolmogorov–Smirnov test, in the Gaussian case, using critical values from simulation. We compute the test statistic for the deciles with 2000 repetitions. For each $\hat{\delta}(\tau_i)$, we simulate the distribution from 5000 samples. The results are presented in Table 4, for 5% level. Regarding the type I error, the results are qualitatively similar to those in Table 1, and the rejection rate is close to 5% in almost all cases. The empirical power is shown in the last two columns for

Table 3
Finite sample power for normal VAR errors ($\alpha = 0.95$).

a_{12}	a_{21}	QAR	CQAR
0.000	0.000	0.110	0.110
−0.300	−0.300	0.082	0.115
	0.000	0.073	0.096
	0.300	0.102	0.098
	0.600	0.163	0.138
0.000	−0.300	0.070	0.098
	0.000	0.069	0.095
	0.300	0.072	0.135
	0.600	0.072	0.141
0.300	−0.300	0.082	0.112
	0.000	0.080	0.147
	0.300	0.041	0.170
	0.600	0.065	0.246
0.600	−0.300	0.161	0.152
	0.000	0.079	0.212
	0.300	0.084	0.316
	0.600	0.067	0.502

Table 4
Simulation finite performance for VAR innovations.

a_{12}	a_{21}	$\alpha = 1$	$\alpha = 0.95$	$\alpha = 0.90$
0.000	0.000	0.065	0.213	0.439
−0.300	−0.300	0.062	0.199	0.396
	0.000	0.058	0.202	0.429
	0.300	0.059	0.173	0.410
	0.600	0.055	0.143	0.397
0.000	−0.300	0.092	0.289	0.549
	0.000	0.063	0.258	0.527
	0.300	0.051	0.246	0.526
	0.600	0.049	0.227	0.532
0.300	−0.300	0.092	0.341	0.672
	0.000	0.058	0.313	0.670
	0.300	0.051	0.319	0.689
	0.600	0.049	0.360	0.694
0.600	−0.300	0.092	0.408	0.744
	0.000	0.062	0.398	0.758
	0.300	0.052	0.465	0.822
	0.600	0.053	0.555	0.855

$\alpha = 0.95$ and $\alpha = 0.90$ respectively. The power gains from inclusion of highly correlated covariates are quite substantial.

In addition, in order to illustrate the available power gains from the inclusion of the covariates in the quantile autoregression model, we derive the asymptotic local power function of the CQAR test and use the asymptotic representation to compare with the CADF and QAR unit root tests. The asymptotic theory is based on the "local-to-unit asymptotics", following Phillips (1987).

Model (1)–(3) contain a unit root under the null hypothesis $H_0 : \alpha = 1$. We allow for local departures from the null hypothesis by setting

$$H_1 : \alpha = 1 - \frac{ca(1)}{n}.$$

The null holds when $c = 0$ and holds "locally" as $n \to \infty$ for $c \neq 0$. The asymptotic theory for near-integrated processes utilizes the Ornstein–Uhlenbeck process, see Cavanagh (1985), Chan and Wei (1987) and Phillips (1987) for more details.[9] The near-integrated limiting form of the test statistic is given in the following proposition, the proof is given in the Appendix.

Proposition 4. *Under Assumptions* A1–A7

$$t_n(\tau) \Rightarrow \delta \frac{\int_0^1 \underline{W}_1^c \mathrm{d}W_1}{\left(\int_0^1 (\underline{W}_1^c)^2\right)^{1/2}} + \sqrt{1-\delta^2} N(0,1) - c\sigma_e \frac{f(F^{-1}(\tau))}{\sqrt{\tau(1-\tau)}} \left(\int_0^1 (\underline{W}_1^c)^2\right)^{1/2}.$$

For the unit root testing problem, large sample power depends on the nuisance parameters δ, σ_e, $f(F^{-1}(\tau))$, and $\sqrt{\tau(1-\tau)}$. The power clearly increases with the fraction $\sigma_e f(F^{-1}(\tau))/\sqrt{\tau(1-\tau)}$, which shifts the distribution to the left. Moreover, δ also affects power by changing the shape of the distribution. Thompson (2004) derives a similar local-to-null asymptotic power functions for robust tests and presents a discussion about how the nuisance parameters change for some distributions and robust estimators.

In order to illustrate the performance of the power functions of CQAR, CADF, and QAR we perform a simple Monte Carlo experiment. We compare the power functions of CQAR, CADF and QAR for standard Normal distribution and t-distribution with 3 degrees of freedom, under trend correction.[10] We generate two independent samples for (u_t) and (x_t) from a given distribution, say Normal with $\sigma_u = \sigma_x = 1$, and t_3, and use a simple model where $e_t = x_t + u_t$. An important feature of this comparison is the set of nuisance parameters. Using the corresponding model, we compute the nuisance parameters necessary to calculate each power function. Important parameters are the standard deviations σ_e and σ_u, and the covariances σ_{eu}, $\sigma_{e\psi}$, and $\sigma_{u\psi}$.[11] Given these parameters, we compute the parameter δ, for the CQAR and QAR, and ρ and R in the CADF. In addition, for given τ, for the CQAR and QAR cases, using the known distribution of (u_t), we can compute $f(F^{-1}(\tau))$, and consequently approximate the power functions for a given parameter c. The power functions were calculated for $c = 1, 2, \ldots, 80$ from simulating 50,000 samples of size 2000 with *i.i.d.* corresponding distribution, and using critical values from Table A in the Appendix.

Fig. 1 displays the power functions in the Normal distribution case. It is possible to notice that the CADF has a higher power than CQAR and QAR when the innovations come from a Gaussian distribution. In addition, there are gains in power in adding covariates in the model, as the comparison between CQAR and QAR curves show.

Fig. 2 shows the power functions for the t_3 case. In the heavier tail distribution case, one can note that the power of CQAR test achieves a major improvement in power relative to CADF and QAR tests. In addition, power of the quantile regression based tests are larger than the CADF test.

5. An application to real exchange rate

An application in which the covariate quantile autoregression test is especially useful is real exchange rates (RER). Models of exchange rate determination rely on the assumption that the purchasing power parity (PPP) hypothesis holds. However, there is conflicting empirical evidence. Recent studies by Papell (1997) and O'Connell (1998) among others, suggest that the issue is not completely solved. In particular, when considering data for the recent flexible rate experience (after 1973), many researchers have

[9] We define $W(\cdot)$ to be standard Brownian motion and define $W^c(\cdot)$ to be the Ornstein–Uhlenbeck process $W^c(r) = \int_0^r \exp\{c(r-s)\}\mathrm{d}W(s)$. In addition, we define the detrended version of W^c as $\underline{W}^c(r) = W^c(r) - \left[\int_0^1 (4-6s)W^c - r\int_0^1 (6-12s)W^c\right]$.

[10] We use median, $\tau = 0.5$, for CQAR and QAR models.

[11] For the normal case it is possible to compute the parameters of interest analytically. They are given by, $\sigma_u = \sigma_{ue} = 1$, $\sigma_{u\psi} = \sigma_{e\psi} = \Phi(0)$, and $\sigma_e = \sqrt{2}$, where $\Phi(\cdot)$ is the distribution function of the standard normal variable. For the t_3 case, we compute these parameters numerically from 50,000 samples of size 2000.

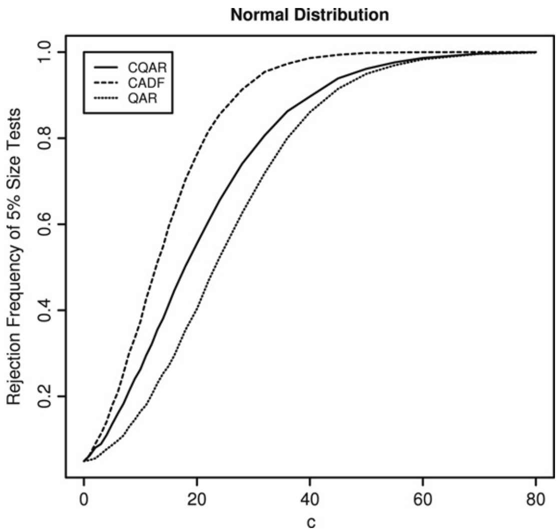

Fig. 1. Power function for normal distribution.

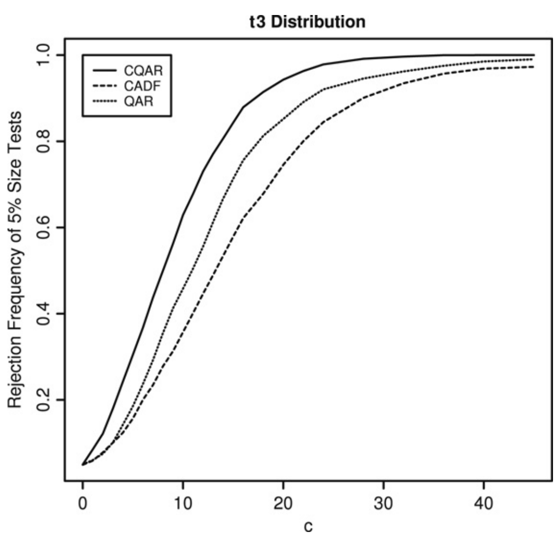

Fig. 2. Power function for t_3 distribution.

been unable to reject the null hypothesis of a unit root (see, e.g. Edison and Pauls (1993)). In addition, several recent works have investigated the possibility of asymmetry in RER time series. Relying on threshold autoregression (TAR) models, Michael et al. (1997) and Bec et al. (2004) found evidence of asymmetry. From the theoretical point of view, as suggested by Obstfeld and Rogoff (2000), asymmetry in RER is natural in the presence of the trading costs for goods.[12]

We apply the proposed CQAR test, estimating Eq. (6), to test whether the real exchange rate has a unit root for Canada, Japan, Switzerland and United Kingdom. The CQAR model is able to convey important insights about dynamics when estimating models for the conditional median and the full range of other conditional quantile functions for the RER. We also apply the CADF test for comparison reasons. The choice of covariates is limited only by the fact that they must be stationary and correlated with the shocks to the real exchange rate. We follow Elliott and Pesavento (2006) in the choice of exogenous covariates. We use the first differences in nominal exchange rates, inflation, money differentials, and interest rates differentials. In order to save space, we present only the results for first differences of nominal exchange rates, although the results for all covariates are essentially the same.[13]

The data employed in this study are monthly data from the IFS database for the period from January of 1973 to January of 2007, covering the post Bretton Woods system and flexible exchange rate regimes. The exchange rates are the end of the month value of the dollar in terms of foreign currencies. In order to choose the appropriate lags of the dependent variables, we used Schwarz information criterion (*SIC*) for CADF, and for covariate quantile autoregression (CQAR) we used a robust Schwarz information criterion based on objective functions defining M-estimators proposed by Machado (1993).[14] In the model selection, Canada and Switzerland do not reject the presence of time trend in the estimated equation, and for Japan and the United Kingdom we do not use time trend in the estimated model.

The results of the CQAR and CADF tests are in Tables 5 and 6, respectively. The results for Canada indicate that it is not possible to reject unit root hypothesis at usual levels of significance using CADF test. The results for the CQAR test show that, when using first difference of nominal exchange rates as covariate, we can reject the unit root hypothesis for the middle deciles. In addition, it is important to note that the behavior of the real exchange rate is not constant, and it is monotonically increasing.

In the case of Japan, the results based on the CADF test indicate that we can reject the null hypothesis of a unit root at 1% level of significance. Moreover, the estimated ρ^2 is very low, indicating that the estimation is very precise. The CQAR results show that it is possible to reject the unit root hypothesis for the first four deciles. Once again, the coefficients have a monotone increasing behavior.

By analyzing the results for Switzerland we can notice that CADF is able to reject the unit root hypothesis at 5% level of significance. Although the evidence for Switzerland is less compelling than for Japan, for which the null hypothesis of a unit root is strongly rejected at the 1% significance level for three deciles, we are able to reject the unit root hypothesis for the first three deciles at the usual levels of significance. Finally, the empirical evidence for the United Kingdom is similar to the one for Canada. It is possible to reject the unit root hypothesis for the second through the sixth deciles.

For Canada and the United Kingdom, it is possible to reject unit root at the middle of the conditional quantile function, and for Japan and Switzerland it is possible to reject unit root for the low part of the conditional quantile function.

We also apply the proposed Kolmogorov–Smirnov test, and calculate the critical values using simulation procedure. We simulate the distribution from 5000 samples. The results are presented in Table 7 and show that it is possible to reject the unit root hypothesis for all countries. For Japan, it is possible to reject the null at 1% level of significance, for Canada and Switzerland it is possible to reject the null at 5% level of significance, and finally, for the United Kingdom, at 10% level of significance.

Therefore, the results show strong evidence that real exchange rates do not behave as a unit root process. In addition, the tests based on covariate quantile autoregression suggest that real exchange rates are not a constant unit root process.

12 It is possible to show that in a two-country stochastic general equilibrium model such costs create a region of no trade where the PPP relationship does not hold. Outside this area, international goods market arbitrage provides an error-correction mechanism that brings the national price level back to equality.

13 The results for all covariates are available upon request.

14 We use the specification of the *SIC* based on the Laplace distribution, which is implemented by an l_1 type of regression.

Table 5
CQAR test for real exchange rates.

τ	Canada			Japan			Switzerland			UK		
	$\hat{\alpha}$	t_n	$\hat{\delta}^2$	$\hat{\alpha}$	t_n	$\hat{\delta}^2$	$\hat{\alpha}$	t_n	$\hat{\delta}^2$	$\hat{\alpha}$	t_n	$\hat{\delta}^2$
0.1	0.995	−1.299	0.046	0.985	−4.515*	0.039	0.994	−2.224***	0.049	0.990	−1.437	0.003
0.2	0.995	−1.256	0.049	0.990	−4.354*	0.011	0.995	−2.529**	0.018	0.992	−2.082***	0.006
0.3	0.996	−1.399	0.024	0.994	−2.831*	0.027	0.997	−2.089***	0.017	0.994	−2.091***	0.019
0.4	0.996	−2.988*	0.040	0.996	−2.301**	0.020	0.997	−1.769	0.029	0.996	−2.571**	0.061
0.5	0.996	−3.174*	0.034	0.998	−1.637	0.052	0.998	−1.469	0.025	0.996	−2.462**	0.034
0.6	0.996	−2.203***	0.024	1.000	−0.133	0.027	0.998	−1.448	0.009	0.996	−2.181**	0.016
0.7	0.997	−2.189***	0.033	1.000	−0.055	0.033	0.999	−0.591	0.009	0.997	−1.874	0.011
0.8	0.997	−1.634	0.009	1.002	1.061	0.005	0.999	−0.571	0.004	0.997	−1.801	0.011
0.9	0.998	−1.099	0.017	1.004	2.019	0.007	0.999	−0.952	0.008	0.997	−1.182	0.011

* Significant at the asymptotic 1% level.
** Significant at the asymptotic 5% level.
*** Significant at the asymptotic 10% level.

Table 6
CADF test for real exchange rates.

Canada			Japan			Switzerland			UK		
$\hat{\alpha}$	t_n	$\hat{\delta}^2$	$\hat{\alpha}$	t_n	$\hat{\delta}^2$	$\hat{\alpha}$	t_n	$\hat{\delta}^2$	$\hat{\alpha}$	t_n	$\hat{\delta}^2$
0.997	−1.591	0.054	0.995	−3.679*	0.044	0.997	−2.494**	0.013	0.994	−2.535**	0.033

* Significant at the asymptotic 1% level.
** Significant at the asymptotic 5% level.

Table 7
Kolmogorov–Smirnov test.

		Canada	Japan	Switzerland	UK
Test statistic		2.942**	4.561*	3.162**	2.837***
Simulated	1%	3.444	3.628	3.626	3.419
Asymptotic critical	5%	2.857	3.121	3.156	3.004
Values	10%	2.660	2.875	2.932	2.807

* Significant at the asymptotic 1% level.
** Significant at the asymptotic 5% level.
*** Significant at the asymptotic 10% level.

6. Conclusion

This paper generalizes the quantile autoregression unit root test to allow stationary covariates variables as well as a linear time trend. The asymptotic theory for the test turns out to be similar to the theory developed by Koenker and Xiao (2004), a weighted combination of Dickey–Fuller and standard normal distributions. However, when allowing for covariates, the weight is determined by the correlation between the equation error and the regression covariates.

Monte Carlo experiments show that the test based on covariate quantile autoregression (CQAR) turns out to be especially advantageous when innovations are non-Gaussian heavy-tailed. In particular, the results show that the quantile autoregression test proposed in this paper presents power gains relative to the QAR test, proposed by Koenker and Xiao (2004), when there is an available stationary covariate and it is included in the estimated model. Furthermore, in the non-Gaussian heavy tailed distribution case, the CQAR test presents more power than the CADF test.

An application to the real exchange rate, using the proposed test, is provided. The evidence based on the point estimates of the autoregressive quantile roots suggests that the real exchange rates series for Canada, Japan, Switzerland and the the United Kingdom are not constant unit root processes.

Acknowledgments

The author would like to thank the Editor and three anonymous referees for their valuable comments and suggestions, and also express his appreciation to Fabio Araujo, Roger Koenker, Luiz Lima, Zhongjun Qu, Zhijie Xiao, and the participants of the seminars at University of Illinois at Urbana-Champaign, Econometrics in Rio, and Robert Bohrer Workshop for helpful comments. All the remaining errors are my own.

Appendix

A.1. Table A: Asymptotic critical values of t-statistic $t_n(\tau)$

δ^2	Demeaned			Detrended		
	1%	5%	10%	1%	5%	10%
1.0	−3.43	−2.86	−2.57	−3.96	−3.41	−3.13
0.9	−3.39	−2.81	2.50	−3.88	−3.33	−3.04
0.8	−3.36	−2.75	−2.46	−3.83	−3.27	−2.97
0.7	−3.30	−2.72	−2.41	−3.76	−3.18	−2.87
0.6	−3.24	−2.64	−2.32	−3.68	−3.10	−2.78
0.5	−3.19	−2.58	−2.25	−3.60	−2.99	−2.67
0.4	−3.14	−2.51	−2.17	−3.49	−2.87	−2.53
0.3	−3.06	−2.40	−2.06	−3.37	−2.73	−2.38
0.2	−2.91	−2.28	−1.92	−3.19	−2.55	−2.20
0.1	−2.78	−2.12	−1.75	−2.97	−2.31	−1.95

A.2. Proofs

Let $\|\cdot\|$ denote the usual Euclidean norm of a vector. In addition to Assumptions A1–A7, in order to ensure tightness, we assume the following regularity condition

B1. There exist a random variable ξ_n and a constant κ_1 ($0 \leq \kappa_1 < 1/2$) such that for all $0 \leq r_1 \leq r_2 \leq 1$, $\sum_{t=[nr_1]}^{[nr_2]} \left\| D_n^{-1} z_t \right\| \leq (r_2 - r_1)\xi_n n^{\kappa_1}$ a.s. In addition, $\sup_n E(\xi_n^{\kappa_2}) \leq C < \infty$ for some $\kappa_2 > 2$.

For $\tau \in \mathcal{T} \equiv [\tau_0, 1-\tau_0]$, $\tau_0 \in (0, 1/2)$, define

$$S_n(\tau, r, b) = n^{-1/2} \sum_{t=1}^{[nr]} \psi_\tau(u_{t\tau}),$$

where $u_{t\tau} = y_t - z_t' b$, and by recentering $\psi_\tau(\cdot)$ at its expectation conditional on z_t, we can define a related quantity $S_n(\tau, r, b)$

$$S_n^d(\tau, r, b) = n^{-1/2} \sum_{t=1}^{[nr]} 1(y_t - z_t' b \le 0) - F(z_t' b).$$

So the following relationship holds

$$S_n(\tau, r, b) = S_n^d(\tau, r, b) + n^{-1/2} \sum_{t=1}^{[nr]} F(z_t' b) - \tau.$$

The next two lemmas show stochastic equicontinuity for the subgradient process. Lemma A1 is an auxiliary result. Along with finite dimensional convergence, it implies weak convergence of $S_n(\tau, r, b)$ on $(D_{[0,1]})^2$.

Lemma A1. *Suppose Assumptions* A1–A7 *and B1 hold, and let $\Phi = [0, 1] \times [0, 1]$ be a parameter set with metric $\rho(\{\tau_1, r_1\}, \{\tau_2, r_2\}) = |\tau_2 - \tau_1| + |r_2 - r_1|$. Then, the process $S_n(\tau, r)$ is stochastically equicontinuous on (Φ, ρ). That is, for any $\epsilon > 0$, $\eta > 0$, there exists a $\delta > 0$ such that for large n,*

$$P\left(\sup_{[\delta]} \|S_n(\tau_1, r_1, \beta(\tau_1)) - S_n(\tau_2, r_2, \beta(\tau_2))\| > \eta\right) < \epsilon,$$

where $[\delta] = \{(s_1, s_2) \in \Phi; s_1 = \{\tau_1, r_1\}, s_2 = \{\tau_2, r_2\}, \rho(s_1, s_2) < \delta\}$.

Proof. For a given τ

$$S_n(\tau, r, \beta(\tau)) = n^{-1/2} \sum_{t=1}^{[nr]} 1(y_t \le z_t' \beta(\tau)) - \tau = n^{-1/2} \sum_{t=1}^{[nr]} 1(F(y_t) \le \tau) - \tau$$

where $F(\cdot)$ is the conditional distribution function of y_t, and the last equality follows because Assumption A7 implies $F(\cdot)$ is absolute continuous and strictly increasing almost everywhere. Define $u_t = F(y_t)$, then u_t has a standard uniform distribution. Hence,

$$S_n(\tau, r, \beta(\tau)) = n^{-1/2} \sum_{t=1}^{[nr]} 1(u_t \le \tau) - \tau.$$

Moreover, $[1(u_t \le \tau) - \tau]$ is a sequence of vector martingale differences by construction. Hence, Bickel and Wichura (1971) (or Lemma A1 of Qu (2008)) applies and the Lemma follows. ■

Lemma A2. *Under Assumptions* A1–A7 *and B1, we have*

$$\sup_{\tau \in \mathcal{T}} \sup_{r \in \mathcal{A}} \sup_{\xi \in D} \left\|S_n^d(\tau, r, \beta(\tau) + D_n^{-1}\xi) - S_n(\tau, r, \beta(\tau))\right\| = o_p(1),$$

where $\mathcal{A} \equiv [c_2, 1 - c_2]$, $c_2 \in (0, 1/2)$ and D is an arbitrary compact set in $\mathbb{R}^{3+p+q}$.

Proof. The proof proceeds along similar lines as that of Lemma 1 of Su and Xiao (in press). Without loss of generality, we can assume the components of z_t are nonnegative. Otherwise, let z_{tk} denote the kth component of z_t, and we can write $z_{tk} = z_{tk}^+ - z_{tk}^- \equiv 1(z_{tk} \ge 0) - 1(z_{tk} < 0)$. Then, z_{tk}^+ and z_{tk}^- are nonnegative and satisfy Assumptions A1–A6. Under the stated assumptions $1(y_t \le z_t'\beta(\tau) + D_n^{-1} z_t' \xi)$ and $F(z_t'\beta(\tau) + D_n^{-1} z_t'\xi)$ are nondecreasing in τ.

Let $N_1 \equiv N_1(n)$ be an integer such that $N_1 = [n^{1/2+d}] + 1$ for some $d \in (0, 1/2)$. We divide the interval $\mathcal{T}$ into N_1 subintervals by points $c_1 = \tau_0 < \tau_1 < \cdots < \tau_{N_1} = 1 - c_1$. The length of each interval is denoted as $\delta^* = (1 - 2c_1)/N_1$. By Assumption A7, for all $\tau_i, \tau_j \in \mathcal{T}$ for such $\left|\tau_i - \tau_j\right| \le \delta^*$, we have: $\left\|\beta(\tau_i) - \beta(\tau_j)\right\| \le (p+q)C_0 \left|\tau_i - \tau_j\right| \le (p+q)C_0\delta^* \equiv C^*$. Suppose that $\tau \in [\tau_{j-1}, \tau_j]$, then

$$\begin{aligned} S_n^d(\tau, r, \beta(\tau) + D_n^{-1}\xi) - S_n(\tau, r, \beta(\tau)) &\le S_n^d(\tau_j, r, \beta(\tau_j) + D_n^{-1}\xi) \\ &- S_n(\tau_{j-1}, r, \beta(\tau_{j-1})) + n^{-1/2} \sum_{t=1}^{[nr]} \{\tau_j - \tau_{j-1}\} \\ &+ n^{-1/2} \sum_{t=1}^{[nr]} \{F(z_t'\beta(\tau_j) + D_n^{-1} z_t'\xi) - F(z_t'\beta(\tau_{j-1}) + D_n^{-1} z_t'\xi)\}. \end{aligned}$$

A reverse inequality holds with $\beta(\tau_j)$ with $\beta(\tau_{j-1})$. Therefore,

$$\begin{aligned} &\left\|S_n^d(\tau, r, \beta(\tau) + D_n^{-1}\xi) - S_n(\tau, r, \beta(\tau))\right\| \\ &\le \left\|S_n^d(\tau_j, r, \beta(\tau_j) + D_n^{-1}\xi) - S_n(\tau_{j-1}, r, \beta(\tau_{j-1}))\right\| \\ &\quad + \left\|S_n^d(\tau_{j-1}, r, \beta(\tau_{j-1}) + D_n^{-1}\xi) - S_n(\tau_j, r, \beta(\tau_j))\right\| \\ &\quad + \left\|n^{-1/2} \sum_{t=1}^{[nr]} \{\tau_j - \tau_{j-1}\}\right\| \\ &\quad + \left\|n^{-1/2} \sum_{t=1}^{[nr]} \{F(z_t'\beta(\tau_j) + D_n^{-1} z_t'\xi) - F(z_t'\beta(\tau_{j-1}) + D_n^{-1} z_t'\xi)\}\right\|. \end{aligned}$$

Therefore, to complete the proof it is sufficient to show that (a), (b), (c), and (d) are $o_p(1)$ uniformly on $\tau \in \mathcal{T}$, $r \in [0, 1]$, and $\xi \in D$.

The term (c) is clearly $o_p(1)$. For the term (d) we follow Qu (2008)

$$\begin{aligned} \sup_{\tau \in \mathcal{T}} \sup_{r \in \mathcal{A}} \sup_{\xi \in D} \|(d)\| &\le \max_{0 \le s \le N_1} \sup_{r \in \mathcal{A}} \sup_{\xi \in D} \|(d)\| \\ &\le \max_{0 \le s \le N_1} \sup_{r \in \mathcal{A}} \sup_{\xi \in D} \left\|n^{-1/2} \sum_{t=1}^{[nr]} \left[f(z_t' b(\tau_j)) - f(z_t' b(\tau_{j-1}))\right] D_n^{-1} z_t' \xi\right\| \\ &\quad + o_p(1) \\ &\le 2 \max_{0 \le s \le N_1} \sup_{r \in \mathcal{A}} \sup_{\xi \in D} \left\|n^{-1/2} \sum_{t=1}^{[nr]} \left[f(z_t' b(\tau_j)) - f(z_t' \beta(\tau_j))\right] D_n^{-1} z_t' \xi\right\| \\ &\quad + \max_{0 \le s \le N_1} \sup_{r \in \mathcal{A}} \sup_{\xi \in D} \left\|n^{-1/2} \sum_{t=1}^{[nr]} \left[f(z_t' \beta(\tau_j)) - f(z_t' \beta(\tau_{j-1}))\right] D_n^{-1} z_t' \xi\right\| \\ &\quad + o_p(1) \end{aligned} \tag{16}$$

where $b(\tau_k)$ ($k = j - 1$ and j) is some vector that lies between $\beta(\tau_k)$ and $\beta(\tau_k) = D_n^{-1}\xi$, and the first inequality follows from a mean value theorem and $\tau_j - \tau_{j-1} \le n^{-1/2-d}$. Now, (16) $= o_p(1)$ if

$$\max_{0 \le s \le N_1} \max_{0 \le t \le n} \left\|f(z_t' b(\tau_j)) - f(z_t' \beta(\tau_j))\right\| = o_p(1) \tag{17}$$

and

$$\max_{0 \le s \le N_1} \max_{0 \le t \le n} \left\|f(z_t' \beta(\tau_j)) - f(z_t' \beta(\tau_{j-1}))\right\| = o_p(1). \tag{18}$$

Eq. (17) holds because $f(s)$ is uniformly continuous in s for all i and

$$\max_{0 \le s \le N_1} \max_{0 \le t \le n} \left\|z_t' b(\tau_j) - z_t' \beta(\tau_j)\right\| = o_p(1)$$

because the vector D_n^{-1} $(\beta(\tau_j) - b(\tau_j)) = O_p(1)$. For (18), note that

$$\begin{aligned} z_t'\beta(\tau_j) - z_t'\beta(\tau_{j-1}) &= \frac{\tau_j - \tau_{j-1}}{f(w_t)} = O_p(\tau_j - \tau_{j-1}) \\ &= O_p(n^{-1/2-d}), \end{aligned} \tag{19}$$

where the first equality follows from the mean value theorem with $z_t'\beta(\tau_{j-1}) \leq w_t \leq z_t'\beta(\tau_j)$, and the second equality follows because $f(\cdot)$ is bounded away from 0. Finally, (19) implies (18) because $f(s)$ is uniformly continuous in s.

Now we turn our attention to terms (a) and (b). We have

$$\begin{aligned}
&\max_{0\leq s\leq N_1} \sup_{r\in\mathcal{A}} \sup_{\xi\in D} (\|(a)\| + \|(b)\|) \\
&\quad \leq 2 \max_{0\leq s\leq N_1} \sup_{r\in\mathcal{A}} \left\| S_n(\tau_j, r, \beta(\tau_j)) - S_n(\tau_{j-1}, r, \beta(\tau_{j-1}))\right\| \\
&\qquad + 2 \max_{0\leq s\leq N_1} \sup_{r\in\mathcal{A}} \sup_{\xi\in D} \left\| S_n^d(\tau_j, r, \beta(\tau_j) + D_n^{-1}\xi) - S_n(\tau_j, r, \beta(\tau_j))\right\|. \quad (20)
\end{aligned}$$

The first term is $o_p(1)$ by Lemma A1. For the second term, because D is compact, for any given $\delta > 0$, D can always be partitioned into a finite number of subsets such that the diameter of each subset is less than or equal to δ. Denote these subsets by $D_1, D_2, \ldots, D_{N(\delta)}$. If $\xi \in D_h (h \in \{1, 2, \ldots, N(\delta)\})$, there exists two points ξ_{h1} and ξ_{h2}, on the boundary of D_h satisfying $z_t'\xi_{h1} \leq z_t'\xi \leq z_t'\xi_{h2}$, leading to

$$\begin{aligned}
&\max_{0\leq s\leq N_1} \sup_{r\in\mathcal{A}} \sup_{\xi\in D} \left\| S_n^d(\tau_j, r, \beta(\tau_j) + D_n^{-1}\xi) - S_n(\tau_j, r, \beta(\tau_j))\right\| \\
&\quad \leq \max_{0\leq s\leq N_1} \sup_{r\in\mathcal{A}} \max_{1\leq h\leq N(\delta)} \max_{k=1,2} \left\| \sum_{t=1}^{[nr]} \{F(z_t'\beta(\tau_j) + D_n^{-1}z_t'\xi_{hk}) - F(z_t'\beta(\tau_j))\}\right\| \\
&\qquad + \max_{0\leq s\leq N_1} \sup_{r\in\mathcal{A}} \max_{1\leq h\leq N(\delta)} \max_{k=1,2} \left\| S_n^d(\tau_j, r, \beta(\tau_j) + D_n^{-1}\xi_{hk}) - S_n(\tau_j, r, \beta(\tau_j))\right\|.
\end{aligned}$$

The first term on the right side is the same order as ξ_{h1} and ξ_{h2}, which can be made arbitrarily small by choosing a small δ. To bound the second term, because $N(\delta)$ and k are finite, it is sufficient to show that for any $\epsilon > 0$, $1 \leq j \leq N_1$, $1 \leq h \leq N(\delta)$ and $k \in \{1, 2\}$,

$$\Pr\left(\max_{0\leq s\leq N_1} \sup_{r\in\mathcal{A}} \left\| S_n^d(\tau_j, r, \beta(\tau_j) + D_n^{-1}\xi_{hk}) - S_n(\tau_j, r, \beta(\tau_j))\right\| > \epsilon\right) \to 0. \quad (21)$$

To this end, let

$$\begin{aligned}
\varsigma_t &= 1(y_t \leq z_t'\beta(\tau_j) + D_n^{-1}z_t'\xi_{hk}) - F(z_t'\beta(\tau_j) + D_n^{-1}z_t'\xi_{hk}) \\
&\quad - 1(y_t \leq z_t'\beta(\tau_j)) + F(z_t'\beta(\tau_j)).
\end{aligned}$$

Therefore, we need to show that $\Pr\left(\sup_{r\in\mathcal{A}} \left\| n^{-1/2}\sum_{t=1}^{[nr]} \varsigma_t\right\| > \epsilon\right) \to 0$. Then ς_i is an array of martingale differences. Apply the Doob inequality

$$\Pr\left(\sup_{r\in\mathcal{A}} \left\| n^{-1/2}\sum_{t=1}^{[nr]} \varsigma_t\right\| > \epsilon/2\right) \leq \frac{16}{n^2\epsilon^4} E\left\|\sum_{t=1}^{n} \varsigma_t\right\|^4.$$

By the Rosenthal inequality

$$\begin{aligned}
T_n &\equiv E\left\|\sum_{t=1}^{n} \varsigma_t\right\|^4 \\
&\leq C\sum_{t=1}^{n} E\left[\|\varsigma_t\|^4\right] + CE\left(\sum_{t=1}^{n} E\|\varsigma_t\|^2\right)^2 \equiv T_{n1} + T_{n2}.
\end{aligned}$$

Clearly, the term $T_{n1} = O(n^{1/2})$. For T_{n2}, by noticing that $E\|\varsigma_t\|^2 \leq |F(z_t'\beta(\tau_j) + D_n^{-1}z_t'\xi_{hk}) - F(z_t'\beta(\tau_j))| \leq C\left\|D_n^{-1}z_t'\xi_{hk}\right\| \leq C\left\|D_n^{-1}z_t'\right\|\|\xi_{hk}\|$, hence by Assumption B1 T_{n2} is bounded such that $T_{n2} \leq CE\left(\sum_{t=1}^{n} CD_n^{-1}z_t'\xi_{hk}\right)^2 = O(n)$. Hence, T_n is $O(n)$, and

$$\Pr\left(\sup_{r\in\mathcal{A}} \left\| n^{-1/2}\sum_{t=1}^{[nr]} \varsigma_t\right\| > \epsilon/2\right) = O(n^{-1}).$$

Consequently,

$$\begin{aligned}
&\Pr\left(\max_{0\leq s\leq N_1} \sup_{r\in\mathcal{A}} \left\| S_n^d(\tau_j, r, \beta(\tau_j) + D_n^{-1}\xi_{hk}) - S_n(\tau_j, r, \beta(\tau_j))\right\| > \epsilon\right) \\
&\quad = O(N_1 n^{-1}) = o(1).
\end{aligned}$$

Therefore, we have shown that (20) is $o_p(1)$ and the lemma follows. ■

Therefore, by finite dimensional convergence and Lemma A2

$$n^{-1/2}\sum_{t=1}^{[nr]} \psi_\tau(u_{t\tau}) \Rightarrow B_\psi^\tau(r) = B_\psi(\tau, r).$$

The finite dimensional convergence to a normal distribution follows from the central limit theorem for a martingale difference sequence.

Proof of Lemma 1. Under the null hypothesis $\alpha = 1$ ($\delta = 0$), $y_t = \mu_1 + \mu_2 t + S_t$, and from Eq. (2) we have $a(L)\Delta S_t = e_t$. Let $k(L) = a(L)^{-1}$. We have $k(L) = k(1) + k^*(1-L)$, where $k^*(L)$ has all roots outside the unit circle because $a^*(L)$ does. Since $\tilde{y}_t = y_t - \mu_1 - \mu_2 t$ from (1) we have $\Delta\tilde{y}_t = \Delta S_t = k(L)e_t$. Therefore, $\frac{1}{\sqrt{n}}\tilde{y}_{[nr]} \Rightarrow k(1)B_e(r) + o_p(1)$. Note that $k(1) = a(1)^{-1}$. In addition, $n^{-1/2}\sum_{t=1}^{[nr]} \psi_\tau(u_{t\tau}) \Rightarrow B_\psi^\tau$ (because $\psi_\tau(u_{t\tau})$ is a measurable function of an α-mixing process). Finally, from Hansen (1992b) Theorem 4.1 $n^{-1}\sum_{t=1}^{n} \tilde{y}_{t-1}\psi_\tau(u_{t\tau}) \Rightarrow a(1)^{-1}\int_0^1 B_e dB_\psi^\tau$.

Hansen (1992b) Theorem 4.1 states that

$$\int_0^1 V_n dV_n' \Rightarrow \int_0^1 BdB' + s\Lambda,$$

where

$$\Lambda_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{t}(U_{ni} - U_{ni-1})w_i - \frac{1}{\sqrt{n}}U_{nt}w_{t+1}. \quad (22)$$

So, we need to show that the remainder Λ_n goes to zero in probability. Under Assumptions A1–A6,

$$\Lambda_n \xrightarrow{p} 0.$$

In order to prove this statement, let $U_{ni} = (n^{-1/2}\sum_1^{i-1} e_i)$ and $w_t = \sum_1^\infty E[\psi(u)|\Im_{t-1}]$. Note that w_t is based on the mixing process $\psi_\tau(u_{t\tau})$ and thus, as in Hansen (1992b, (A3))

$$n^{-1/2}\sup_{i\leq n} |w_i| \xrightarrow{p} 0. \quad (23)$$

Looking at the last term of (22) first note that

$$|n^{-1/2}U_{nn}w_{n+1}| \leq |U_{nn}| \cdot |n^{-1/2}w_{n+1}|.$$

Under Assumptions A1–A6, the partial sum process of $\{e_t\}$ converges and the first term is $O_p(1)$, while the second term converges to zero in probability by (23). Thus the product vanishes asymptotically. As to the summation term in (22) notice that

$$U_{ni} - U_{ni-1} = n^{-1/2}e_i.$$

Hence

$$n^{-1/2}\sum_{i=1}^{n}(U_{ni} - U_{ni-1})w_i = n^{-1}\sum_{i=1}^{n} e_i w_i.$$

Applying Holder's inequality followed by the weak convergence of $\{e_t\}$ and (23), this term also converges to zero in probability. Thus the lemma follows. ■

We need to consider the limiting distribution of

$$\begin{bmatrix} n^{-1/2}\sum_t \Delta y_{t-1}\psi_\tau(u_{t\tau}) \\ \vdots \\ n^{-1/2}\sum_t \Delta y_{t-p}\psi_\tau(u_{t\tau}) \\ n^{-1/2}\sum_t x_{t-q_1}\psi_\tau(u_{t\tau}) \\ \vdots \\ n^{-1/2}\sum_t x_{t+q_2}\psi_\tau(u_{t\tau}) \end{bmatrix}. \tag{24}$$

By assumption $E(\phi_t\phi_t') > 0$, where $\phi_t = (\Delta y_{t-1}, \ldots, \Delta y_{t-q}, x_{t-q_1}', \ldots, x_{t+q_2}')'$. Because Assumption A1 effectively implies that the regression error u_t is orthogonal to the lagged differences of the dependent variable $\Delta y_{t-1}, \ldots, \Delta y_{t-p}$ and the leads and lags of the stationary covariates x_t, and $E(\phi_t\psi_\tau(u_{t\tau})) = 0$, one can be shown that, as a function of τ, (24) converges to a $(p+q)$-dimensional Brownian Bridge, B_ψ^τ, with elements in the covariance matrix given by $E(\phi_t\phi_t')$. It is important to note that, for given τ, this is a normal variate $\Phi_\tau = [\Phi_1, \ldots, \Phi_{p+q}]'$ with covariance matrix $\tau(1-\tau)\Omega_\Phi$ where the elements of Ω_Φ are the elements of the matrix $E(\phi_t\phi_t')$. Following Koenker and Xiao (2004) we use the latter notation in the results.

Proof of Lemma 2. Using the previous lemma and noting that $n^{-3/2}\sum_{t=1}^n r\psi_\tau(u_{t\tau}) \Rightarrow \int r\mathrm{d}B_\psi^\tau$, and $n^{-1/2}\sum_{t=1}^n \psi_\tau(u_{t\tau}) \Rightarrow \int \mathrm{d}B_\psi^\tau$, as argued in the proof of the above lemma. Now, notice that $\phi_t \in \Im_{t-1}$ and $E(\psi_\tau(u_{t\tau})|\Im_{t-1}) = 0$, $\phi_t\psi_\tau(u_{t\tau})$ is a martingale difference sequence and thus $n^{-1/2}\sum_{t=1}^n \phi_t\psi_\tau(u_{t\tau})$ satisfies a central limit theorem. By Lemma A2 the autoregression process is tight and thus the limiting variate viewed as a random function of τ, is a Brownian bridge over $\tau \in \mathcal{T}$, and thus $n^{-1/2}\sum_{t=1}^n \phi_t\psi_\tau(u_{t\tau}) \Rightarrow \Omega_\Phi^{1/2}B_\psi^\tau$. Hence, for fixed τ, $n^{-1/2}\sum_{t=1}^n \phi_t\psi_\tau(u_{t\tau})$ converges to $(p+q)$-dimensional vector normal variate with covariance matrix $\tau(1-\tau)\Omega_\Phi$. Therefore, the lemma follows. ■

It is important to note that in Lemma 2, Φ_τ is independent with $\int_0^1 \bar{B}_e\mathrm{d}B_\psi^\tau$.

Proof of Lemma 3. Koenker and Xiao (2004) proved the result for the estimated equation without covariates and a linear time trend. In order to extend their results, we observe that in the Koenker and Xiao (2004) proof of Theorem 2.2 (pages 784–786) it is necessary to incorporate the time trend regressor and the covariates. So, first write $\sum_{t=1}^n \int_0^{(D_n^{-1}v)'\tilde{z}_t}\{I(u_{t\tau} \le s) - I(u_{t\tau} < 0)\}\mathrm{d}s$ as

$$\sum_{t=1}^n (u_{t\tau} - (D_n^{-1}v)'\tilde{z}_t)\{I(0 > u_{t\tau} > (D_n^{-1}v)'\tilde{z}_t) - I(0 < u_{t\tau} < (D_n^{-1}v)'\tilde{z}_t)\}.$$

Now we consider the limit of

$$\sum_{t=1}^n (u_{t\tau} - v'D_n^{-1}\tilde{z}_t)I(0 < u_{t\tau} < v'D_n^{-1}\tilde{z}_t).$$

Denote

$$U_n(v) = \sum_{t=1}^n w_t(v) \quad \text{where } w_t(v) = (v'D_n^{-1}\tilde{z}_t - u_{t\tau})I(0 < u_{t\tau} < v'D_n^{-1}\tilde{z}_t).$$

Following Knight (1989), consider truncation of $v_2 n^{-1/2}\tilde{y}_{t-1}$ at some finite number $m > 0$ and denote

$$U_{nm} = \sum_{t=1}^n w_{tm}(v),$$

$$w_{tm}(v) = (v'D_n^{-1}\tilde{z}_t - u_{t\tau})I(0 < u_{t\tau} < v'D_n^{-1}\tilde{z}_t)M_t$$

$$M_t = I(0 < v_2 n^{-1/2}\tilde{y}_{t-1} < m).$$

Further define

$$\bar{w}_{tm}(v) = E\left\{(v'D_n^{-1}\tilde{z}_t - u_{t\tau})I(0 < u_{t\tau} < v'D_n^{-1}\tilde{z}_t)M_t|\Im_{t-1}\right\},$$

and

$$\bar{U}_{nm}(v) = \sum_{t=1}^n \bar{w}_{tm}(v),$$

then $\{w_{tm}(v) - \bar{w}_{tm}(v)\}$ is a martingale difference sequence. In addition,

$$\begin{aligned}
\bar{U}_{nm}(v) &= \sum_{t=1}^n E\left\{(v'D_n^{-1}\tilde{z}_t - u_{t\tau})I(0 < u_{t\tau} < v'D_n^{-1}\tilde{z}_t)M_t|\Im_{t-1}\right\} \\
&= \sum_{t=1}^n \int_{F_t^{-1}(\tau)}^{[(n^{-1/2}v)'\tilde{z}_t + F_t^{-1}(\tau)]M_t} \left[(v'D_n^{-1}\tilde{z}_t + F_t^{-1}(\tau))M_t - r\right] f_t(r)\mathrm{d}r \\
&= \sum_{t=1}^n \int_{F_t^{-1}(\tau)}^{[(n^{-1/2}v)'\tilde{z}_t + F_t^{-1}(\tau)]M_t} \left[\int_r^{[v'D_n^{-1}\tilde{z}_t + F_t^{-1}(\tau)]M_t} \mathrm{d}s\right] f_t(r)\mathrm{d}r \\
&= \sum_{t=1}^n \int_{F_t^{-1}(\tau)}^{[(n^{-1/2}v)'\tilde{z}_t + F_t^{-1}(\tau)]M_t} \left[\int_{F_t^{-1}(\tau)}^{s} f_t(r)\mathrm{d}r\right] \mathrm{d}s \\
&= \sum_{t=1}^n \int_{F_t^{-1}(\tau)}^{[(n^{-1/2}v)'\tilde{z}_t + F_t^{-1}(\tau)]M_t} [s - F_t^{-1}(\tau)] \\
&\quad \times \left[\frac{F_t(s) - F_t(F_t^{-1}(\tau))}{s - F_t^{-1}(\tau)}\right]\mathrm{d}s.
\end{aligned}$$

Under Assumption A7, by uniform integrability of $\{f_t(x_n)\}$ for $x_n \to 0$ and the definition of $w_{tm}(v)$, we can write,

$$\begin{aligned}
\bar{U}_{nm}(v) &= \sum_{t=1}^n \int_{F_t^{-1}(\tau)}^{[(n^{-1/2}v)'\tilde{z}_t + F_t^{-1}(\tau)]M_t} [s - F_t^{-1}(\tau)]f_t(r) \\
&\quad \times [F_t^{-1}(\tau)]\mathrm{d}s + o_p(1). \\
&= \frac{1}{2}\sum_{t=1}^n f_t[F^{-1}(\tau)]v'[D_n^{-1}\tilde{z}_t\tilde{z}_t'D_n^{-1}]vM_t + o_p(1),
\end{aligned}$$

which in turn, by A7 leads to

$$\bar{U}_{nm}(v) \Rightarrow \eta_m = \frac{1}{2}f(F^{-1}(\tau))v'\Psi_{1m}v$$

where we denote $\Psi_{1m} = \begin{bmatrix} \int_0^1 \bar{B}_e\bar{B}_e' I(0 \le v_2'B_e(s) \le m) & 0_{(p+q)}' \\ 0_{(p+q)} & \Omega_\Phi \end{bmatrix}$.

Following Pollard (1984, p171) and noticing that $(v'D_n^{-1}\tilde{z}_t)I(0 < v_2 n^{-1/2}y_{t-1}^* < m) \xrightarrow{p} 0$ uniformly in t,

$$\sum_{t=1}^n E\left[w_{tm}(v)^2|\Im_{t-1}\right] \le \max\left\{(v'D_n^{-1}\tilde{z}_t)I(0 < v_2 n^{-1/2}\tilde{y}_{t-1} < m)\right\} \times \sum_{t=1}^n \bar{w}_{tm}(v) \xrightarrow{p} 0.$$

Thus the following summation of martingale difference

$$\sum_{t=1}^n \{w_{tm}(v) - \bar{w}_{tm}(v)\}$$

converges to zero in probability and the limiting distribution of $\sum_{t=1}^n w_{tm}(v)$ is the same as that of $\sum_{t=1}^n \bar{w}_{tm}(v)$, that is,

$$U_{nm}(v) \Rightarrow \eta_m.$$

Let $m \to \infty$ we have,

$$\eta_m \Rightarrow \eta = \frac{1}{2} f(F^{-1}(\tau)) v' \Psi_1 v I(v_2 B_e(s) > 0)$$

and $\Psi_1 = \begin{bmatrix} \int_0^1 \bar{B}_e \bar{B}_e' I(0 \leq v_2' B_e(s)) & 0'_{(p+q)} \\ 0_{(p+q)} & \Omega_\Phi \end{bmatrix}$.

Finally, by a similar argument as Herce (1996) one can show that

$$\lim_{m\to\infty} \limsup \Pr[|U_n(v) - U_{nm}(v)| \geq \varepsilon] = 0.$$

Similarly,

$$\sum_{t=1}^{n} (u_{t\tau} - (D_n^{-1} v)' \tilde{z}_t) I(0 > u_{t\tau} > (D_n^{-1} v)' \tilde{z}_t) \Rightarrow \frac{1}{2} f(F^{-1}(\tau)) v' \Psi_2 v$$

where $\Psi_2 = \begin{bmatrix} \int_0^1 \bar{B}_e \bar{B}_e' I(v_2' B_e(s) \leq 0) & 0'_{(p+q)} \\ 0_{(p+q)} & \Omega_\Phi \end{bmatrix}$. Therefore,

$$\sum_{t=1}^{n} (u_{t\tau} - (D_n^{-1} v)' z_t^*) \{ I(0 > u_{t\tau} > (D_n^{-1} v)' \tilde{z}_t)$$
$$- I(0 < u_{t\tau} < (D_n^{-1} v)' \tilde{z}_t) \} \Rightarrow \frac{1}{2} f(F^{-1}(\tau)) v' \Psi v$$

where $\Psi = \begin{bmatrix} \int_0^1 \bar{B}_e \bar{B}_e' & 0'_{(p+q)} \\ 0_{(p+q)} & \Omega_\Phi \end{bmatrix}$. ■

Proof of Proposition 1. Using the identity of Knight (1989), Eq. (10) can be written as

$$\sum_{t=1}^{n} [\rho_\tau(u_{t\tau} - (D_n^{-1} v)' \tilde{z}_t) - \rho_\tau(u_{t\tau})]$$
$$= -\sum_{t=1}^{n} v' D_n^{-1} \tilde{z}_t \psi_\tau(u_{t\tau}) + \sum_{t=1}^{n} (u_{t\tau} - v' D_n^{-1} \tilde{z}_t)$$
$$\times \{ I(0 > u_{t\tau} > v' D_n^{-1} \tilde{z}_t) - I(0 < u_{t\tau} < v' D_n^{-1} \tilde{z}_t) \}.$$

By Lemmas 2 and 3,

$$\begin{aligned} H_n(v) &= \sum_{t=1}^{n} [\rho_\tau(u_{t\tau} - (D_n^{-1} v)' \tilde{z}_t) - \rho_\tau(u_{t\tau})] \\ &\Rightarrow -v' \Phi_\tau^* + \frac{1}{2} f(F^{-1}(\tau)) v' \Psi v \\ &:= H(v) \end{aligned}$$

and the finite dimensional distributions of $H_n(\cdot)$ converge weakly to those of $H(\cdot)$. By convexity Lemma of Pollard (1991) and the arguments of Knight (1989), notice that $H_n(v)$ and $H(v)$ are minimized at $\hat{v} = D_n(\hat{\beta}(\tau) - \beta(\tau))$ and

$$\frac{1}{f(F^{-1}(\tau))} \begin{bmatrix} \int_0^1 \bar{B}_e \bar{B}_e' & 0_{3\times(p+q)} \\ 0_{(p+q)\times 3} & \Omega_\Phi \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 \bar{B}_e dB_\psi^\tau \\ \Phi \end{bmatrix},$$

respectively, and by the Lemma A of Knight (1989) we have,

$$D_n(\hat{\beta}(\tau) - \beta(\tau)) \Rightarrow \frac{1}{f(F^{-1}(\tau))} \begin{bmatrix} \int_0^1 \bar{B}_e \bar{B}_e' & 0_{3\times(p+q)} \\ 0_{(p+q)\times 3} & \Omega_\Phi \end{bmatrix}^{-1}$$
$$\times \begin{bmatrix} \int_0^1 \bar{B}_e dB_\psi^\tau \\ \Phi_\tau \end{bmatrix}. \quad ■$$

Proof of Corollary 1. First, note that $\begin{bmatrix} \int_0^1 \bar{B}_e \bar{B}_e' & 0_{3\times(p+q)} \\ 0_{(p+q)\times 3} & \Omega_\Phi \end{bmatrix}$ is a block diagonal matrix. Using the definition $\bar{B}_e(r) = [1, r, a(1)^{-1} B_e(r)]'$, we take the inverse of the stated matrix and multiply the result by $\begin{bmatrix} \int_0^1 \bar{B}_e dB_\psi^\tau \\ \Phi_\tau \end{bmatrix}$. Now using the vector $e_z = (0, 0, 1, 0, \ldots, 0)$, we pre-multiply $e_z \begin{bmatrix} \left(\int_0^1 \bar{B}_e \bar{B}_e'\right)^{-1} & 0_{3\times(p+q)} \\ 0_{(p+q)\times 3} & \Omega_\Phi^{-1} \end{bmatrix} \begin{bmatrix} \int_0^1 \bar{B}_e dB_\psi^\tau \\ \Phi_\tau \end{bmatrix}$ and the corollary follows. ■

Proof of Proposition 2. First, note that we can rewrite (13) as

$$t_n(\tau) = \frac{\widehat{f(F^{-1}(\tau))}}{\sqrt{\tau(1-\tau)}} \left(\frac{1}{n^2} Y_{-1}^T M_z Y_{-1} \right)^{1/2} n(\hat{\alpha}(\tau) - 1).$$

Note that the term $\frac{1}{n^2} Y_{-1}^T M_z Y_{-1} = \frac{1}{n^2} Y_{-1}^T M_z M_z Y_{-1}$ and we can rewrite it as $n^{-2} \sum (w_t - \hat{w}_t)^2$ where $\hat{w}_t = z(z'z)^{-1} z' w_t$ and $w_t = \tilde{y}_{t-1}$. Now by continuous mapping theorem,

$$n^{-2} \sum (w_t - \hat{w}_t)^2 \Rightarrow a(1)^{-2} \int_0^1 \left(B_e(r) - \left[\int_0^1 (4 - 6r) B_e \right.\right.$$
$$\left.\left. - r \int_0^1 (6 - 12r) B_e \right] \right)^2$$

and this is asymptotically equal to $a(1)^{-2} \int_0^1 \underline{B}_e^2$, where $\underline{B}_e$ is the detrended Brownian motion. Notice that the term $\frac{1}{n^2} Y_{-1}^T M_z M_z Y_{-1}$ is invariant with respect to the constant and time trend in y_{t-1}, since M_Z projects onto the space orthogonal to $Z = (1, t, \Delta y_{t-1}, \ldots, \Delta y_{t-p}, x_{t-q_2}, \ldots, x_{t+q_1})$, implying that we can define $Y_{-1} = y_{t-1}$ as well as $Y_{-1} = \tilde{y}_{t-1}$.

Finally, since $\widehat{f(F^{-1}(\tau))}$ is a consistent estimator for $f(F^{-1}(\tau))$, by Slutsky theorem, and the continuous mapping theorem and Corollary 1

$$t_n(\tau) \Rightarrow t(\tau) = \frac{1}{\sqrt{\tau(1-\tau)}} \left[\int_0^1 \underline{B}_e^2 \right]^{-1/2} \left[\int_0^1 \underline{B}_e dB_\psi^\tau \right]. \quad ■$$

The following Lemma provides asymptotic theory for near-integrated process and it is an auxiliary result to derive the Power Function for CQAR test.

Lemma 4. *Under Assumptions A1–A7*

$$n(\hat{\alpha} - \alpha) \Rightarrow \frac{a(1)}{f(F^{-1}(\tau))} \left[\int_0^1 \left(\underline{B}_e^c\right)^2 \right]^{-1} \left[\int_0^1 \underline{B}_e^c dB_\psi^\tau \right].$$

Proof. In the same lines of Lemma 1, but using near-integrated asymptotic theory, by Theorem 4.4 part (a) in Hansen (1995) $n^{-1/2} \tilde{y}_{[nr]} \Rightarrow a(1)^{-1} B_e^c(r)$. In addition, $n^{-1/2} \sum_{t=1}^{[nr]} \psi_\tau(u_{t\tau}) \Rightarrow B_\psi^\tau$. Finally, from Theorem 4.4 part (b) of Hansen (1992b) $n^{-1} \sum_{t=1}^{n} \tilde{y}_{t-1} \psi_\tau(u_{t\tau}) \Rightarrow a(1)^{-1} \int_0^1 B_e^c dB_\psi^\tau$. Therefore, using the exact same technique as in proof of Proposition 1

$$n(\hat{\alpha} - \alpha) \Rightarrow \frac{a(1)}{f(F^{-1}(\tau))} \left[\int_0^1 \left(\underline{B}_e^c\right)^2 \right]^{-1} \left[\int_0^1 \underline{B}_e^c dB_\psi^\tau \right]$$

where $\underline{B}_e^c = B_e^c - \left[\int_0^1 (4 - 6r) B_e^c - r \int_0^1 (6 - 12r) B_e^c \right]$. ■

Proof of Proposition 4.

$$t_n(\tau) = \frac{\widehat{f(F^{-1}(\tau))}}{\sqrt{\tau(1-\tau)}} \left(\frac{1}{n^2} Y_{-1}^T M_z Y_{-1} \right)^{1/2} n(\hat{\alpha}(\tau) - 1)$$

under local departures of the null hypothesis, $\alpha = 1 - \frac{ca(1)}{n}$, then

$$t_n(\tau) = \frac{f(\widehat{F^{-1}(\tau)})}{\sqrt{\tau(1-\tau)}} \left(\frac{1}{n^2} Y_{-1}^T M_z Y_{-1}\right)^{1/2} n\left(\hat{\alpha}(\tau) - \alpha\right) - ca(1) \frac{f(\widehat{F^{-1}(\tau)})}{\sqrt{\tau(1-\tau)}} \left(\frac{1}{n^2} Y_{-1}^T M_z Y_{-1}\right)^{1/2}.$$

Lemma 4, consistency of $f(\widehat{F^{-1}(\tau)})$, convergence of $\left(\frac{1}{n^2} Y_{-1}^T M_z Y_{-1}\right)$, and the exact same decomposition as in Section 3.1 yield

$$t_n(\tau) \Rightarrow \delta \frac{\int_0^1 \underline{W}_1^c \mathrm{d}W_1}{\left(\int_0^1 (\underline{W}_1^c)^2\right)^{1/2}} + \sqrt{1-\delta^2} N(0, 1) - c\sigma_e \frac{f(F^{-1}(\tau))}{\sqrt{\tau(1-\tau)}} \left(\int_0^1 (\underline{W}_1^c)^2\right)^{1/2}. \quad \blacksquare$$

References

Bassett, G.W., Koenker, R., 1982. An empirical quantile function for linear models with iid errors. Journal of the American Statistical Association 77, 407–415.

Bec, F., Salem, M.B., Carrasco, M., 2004. Tests for unit root versus threshold specification with an application to the PPP. Journal of Business and Economic Statistics 22, 382–395.

Bickel, P.J., Wichura, M.J., 1971. Convergence criteria for multiparameter stochastic processes and some applications. The Annals of Mathematical Statistics 42, 1656–1670.

Bofinger, E., 1975. Estimation of a density function using order statistics. Australian Journal of Statistics 17, 1–7.

Cavanagh, C.L., 1985. Roots local to unity. Harvard University, Unpublished Manuscript.

Chan, N.H., Wei, C.Z., 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Annals of Statistics 15, 1050–1063.

Edison, H., Pauls, B., 1993. A re-assessment of the relationship between real exchange rates and real interest rates: 1974–1990. Journal of Monetary Economics 31, 165–187.

Elliott, G., Jansson, M., 2003. Testing for unit root with stationary covariates. Journal of Econometrics 115, 75–89.

Elliott, G., Pesavento, E., 2006. On the failure of purchasing power parity for bilateral exchange rates after 1973. Journal of Money, Credit, and Banking 38, 1405–1430.

Hansen, B.E., 1992a. Consistent covariance matrix estimation for dependent heterogeneous processes. Econometrica 60, 967–972.

Hansen, B.E., 1992b. Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 489–500.

Hansen, B.E., 1995. Rethinking the univariate approach to unit root testing: Using covariates to increase power. Econometric Theory 11, 1148–1171.

Hasan, M.N., 2001. Rank tests of unit root hypothesis with infinite variance errors. Journal of Econometrics 104, 49–65.

Hasan, M.N., Koenker, R., 1997. Robust rank tests of the unit root hypothesis. Econometrica 65, 133–161.

Herce, M.A., 1996. Asymptotic theory of lad estimation in a unit root process with finite variance errors. Econometric Theory 12, 129–153.

Knight, K., 1989. Limit theory for autoregressive parameter estimates in an infinite variance random walk. The Canadian Journal of Statistics 17, 261–278.

Knight, K., 1991. Limiting theory for M estimates in an integrated infinite variance process. Econometric Theory 7, 200–212.

Koenker, R., 2005. Quantile Regression. Cambridge University Press, New York, NY.

Koenker, R., Bassett, G.W., 1978. Regression quantiles. Econometrica 46, 33–49.

Koenker, R., Xiao, Z., 2004. Unit root quantile autoregression inference. Journal of the American Statistical Association 99, 775–787.

Lucas, A., 1995. Unit root tests based on M estimators. Econometric Theory 11, 331–346.

Machado, J.A.F., 1993. Robust model selection and M-estimation. Econometric Theory 9, 478–493.

Michael, P., Nobay, A., Peel, D., 1997. Transaction costs and nonlinear adjustment in real exchange rates: An empirical investigation. Journal of Political Economy 105, 862–880.

Nelson, C.R., Plosser, C.I., 1982. Trends and random walks in macroeconomic time series: Some evidence and implications. Journal of Monetary Economics 10, 139–162.

Obstfeld, M., Rogoff, K., 2000. The six major puzzles in international macroeconomics: Is there a common cause? NBER Working Paper 7777.

O'Connell, P.G., 1998. The overvaluation of purchasing power parity. Journal of International Economics 44, 1–20.

Papell, D.H., 1997. Searching for stationarity: Purchasing power parity under the current float. Journal of International Economics 43, 313–323.

Park, J.Y., Phillips, P.C.B., 1988. Statistical inference in regression with integrated process: Part 1. Econometric Theory 4, 468–497.

Pesavento, E., 2007. Near-optimal unit root test with stationary covariate with better finite sample size. Mimeo, Emory University.

Phillips, P.C.B., 1987. Toward a unified asymptotic theory for autoregression. Biometrika 74, 535–574.

Phillips, P.C.B., 1995. Fully modified least squares and vector autoregression. Econometrica 63, 1023–1078.

Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regression with I(1) processes. Review of Economic Studies 57, 99–125.

Pollard, D., 1984. Convergence of stochastic processes. Springer-Verlag, New York.

Pollard, D., 1991. Asymptotics for least absolute deviation regression estimators. Econometric Theory 7, 186–199.

Qu, Z., 2008. Testing for structural change in regression quantiles. Journal of Econometrics 146, 170–184.

Rothenberg, T.J., Stock, J.H., 1997. Inference in a nearly integrated autoregressive model with nonnormal innovations. Journal of Econometrics 80, 269–286.

Seo, B., 1999. Distribution theory for unit root tests with conditional heteroskedasticity. Journal of Econometrics 91, 113–144.

Shin, D.W., So, B.S., 1999. Unit root tests based on adaptive maximum likelihood estimation. Econometric Theory 15, 1–23.

Su, L., Xiao, Z., 2008. Testing for parameter stability in quantile regression models. Statistics and Probability Letters 78, 2768–2775.

Thompson, S.B., 2004. Robust tests of the unit root hypothesis should not be 'modified'. Econometric Theory 20, 360–381.