

Quantile Cointegration in the Autoregressive Distributed-Lag Modelling Framework*

JIN SEO CHO

School of Economics

Yonsei University

50 Yonsei-ro, Seodaemun-gu, Seoul, 120-749, Korea

jinseocho@yonsei.ac.kr

TAE-HWAN KIM

School of Economics

Yonsei University

50 Yonsei-ro, Seodaemun-gu, Seoul, 120-749, Korea

tae-hwan.kim@yonsei.ac.kr

YONGCHEOL SHIN[†]

Department of Economics and Related Studies

University of York

Heslington, York, YO10 5DD, U.K.

yongcheol.shin@york.ac.uk

May 25, 2015

Abstract

Xiao (2009) develops a novel estimation technique for quantile cointegrated time series by extending Phillips and Hansen's (1990) semiparametric approach and Saikkonen's (1991) parametrically augmented approach. This paper extends Pesaran and Shin's (1998) autoregressive distributed-lag approach into quantile regression by jointly analysing short-run dynamics and long-run cointegrating relationships across a range of quantiles. We derive the asymptotic theory and provide a general package in which the model can be estimated and tested within and across quantiles. We further affirm our theoretical results by Monte Carlo simulations. The main utilities of this analysis are demonstrated through the empirical application to the dividend policy in the U.S.

Keywords: QARDL, Quantile Regression, Long-run Cointegrating Relationship, Dividend Smoothing, Time-varying Rolling Estimation.

JEL classifications: C22, G35.

*All of the mathematical proofs and the GAUSS codes for the simulations and the empirical estimations are available at <http://web.yonsei.ac.kr/jinseocho/research.htm>.

[†]Corresponding Author: tel: +44 (0)19 043 23757. fax: +44 (0)19 043 23759.

1 Introduction

The statistical properties of cointegrated time series have been extensively investigated over the last three decades. The fully modified OLS estimator of Phillips and Hansen (1990) has been the most widely used method in linear time series modelling. However, this semiparametric approach is static and shows only the long-run relationship between integrated series, leaving the associated short-run dynamics to be recovered through a potentially inefficient two-step approach, as in Engle and Granger (1987). In response, Pesaran and Shin (1998) propose the autoregressive distributed lag error correction model (ARDL-ECM) that allows us to simultaneously investigate the long-run relationship and the short-run dynamics in a fully parametric manner. In particular, they generalise the ARDL approach for cointegration and develop the asymptotic theory for estimation and inference. Furthermore, Pesaran, Shin, and Smith (2001) develop a pragmatic bounds-testing procedure for the existence of a stable long-run relationship, which is valid irrespective of whether the underlying regressors are either $I(1)$, mutually cointegrated, or $I(0)$.¹

Recently, the literature on quantile time series regression has been rapidly growing, *e.g.*, Koenker and Xiao (2004, 2006).² Importantly, Xiao (2009) advances a quantile cointegration approach in a static regression and develops the semiparametric fully modified and the parametrically augmented quantile estimators, which can be regarded as the quantile counterparts of the estimators proposed by Phillips and Hansen (1990) and Saikkonen (1991). It is well established that the quantile estimator is consistent and asymptotically normal when the stationarity condition is satisfied together with other regularity conditions, *e.g.*, Koenker and Basset (1978), Phillips (1991), Koenker and Zhao (1996), and Kim and White (2003). In this regard, Xiao's approach is a challenging contribution to the quantile regression context with nonstationary variables. As a pioneer of cointegration analysis, Granger (2010) provides further insightful discussions on the analysis of possibly cointegrated quantile time series. Xiao's (2009) approach has also been adopted by a number of studies, documenting evidence that the conventional cointegration analysis focusing on the mean behavior may not be sufficiently informative.³

¹The flexibility and utility of the ARDL technique are reflected in the vast literature that adopts its applications for the analysis of a wide range of economic variables (*e.g.*, the video, available at <http://www.youtube.com/watch?v=d9E8BKsosis>, which demonstrates its applications in Microfit and Eviews).

²Quantile regression models have been widely used in a number of fields, notably the analysis of stock market returns (*e.g.*, Barnes and Huhes, 2002) and in labour economics (*e.g.*, Martins and Pereira, 2004). They have also become an important tool in risk management; see Engle and Manganelli (2004) for the value at risk (VaR) evaluation; Adrian and Brunnermeier (2010) for conditional VaR; and Acharya *et al.* (2011) for marginal expected shortfall.

³Lee and Zeng (2011) find that the response of spot oil prices to shocks in one-month futures is much greater with higher spot prices than with lower prices. Utilising the post Asian financial crisis sample of 1999–2010, Burdekin and Siklos (2012) document evidence of integration of the Shanghai stock market with the US and many regional stock markets, though cointegration is found to be prevalent at the higher end of the distribution. Tsong and Lee (2013) examine an empirical validity of the Fisher hypothesis for six OECD countries and find that the cointegrating relationships between nominal interest rate and inflation display location

In this paper, we aim to contribute to this growing literature by proposing the dynamic quantile ARDL-ECM (QARDL-ECM), in which we can simultaneously address both the long-run (cointegrating) relationship and the associated short-run dynamics across a range of quantiles in a fully parametric setting. As this is an extension of the ARDL-ECM developed by Pesaran and Shin (1998) into the quantile regression context, it is expected that all of the optimal estimation properties can be obtained in a similar manner to Xiao's (2009) quantile extension of Phillips and Hansen (1990) and Saikkonen (1991).

We provide an asymptotic theory for estimating and testing the QARDL model with nonstationary regressors. The QARDL estimators of the short-run dynamic parameters and the long-run cointegrating parameters are shown to asymptotically follow the (mixture) normal distribution. We also show that the null distribution of the Wald statistics for testing the restrictions on the short- and long-run parameters within and across quantiles weakly converges to a chi-squared distribution. Via Monte Carlo simulation studies, we find that the overall simulation results, focusing on empirical size and power of the statistics, provide strong support for theoretical predictions both in the case with fixed QARDL orders and in the case where (unknown) QARDL orders are consistently selected by the Bayesian information criterion (BIC).

In a seminal study on dividend policy, Lintner (1956) observes that firms gradually adjust dividends in response to changes in earnings toward the long-run target payout ratio. Empirical researches at both firm and aggregate levels generally support Lintner's partial adjustment framework, *e.g.*, Fama and Babiak (1968) and Marsh and Merton (1987). Recently, Brav, Graham, Harvey and Michaely (2005) surveyed 384 financial executives to determine the factors that drive dividend and share repurchase decisions. Although the new survey evidence is mostly consistent with Lintner's observations, the link between dividends and earnings was substantially weak. Using the firm-level data in the US, Leary and Michaely (2011) document that dividend smoothing has steadily increased over the past century, even before firms began using share repurchases in the mid-1980s. Chen, Da, and Priestley (2012) also demonstrate that aggregate dividends are dramatically more smoothed in the postwar period than in the prewar period.

However, all of these studies examine dividend behavior only at the conditional mean and do not investigate an important possibility that the dividend policy may be fundamentally heterogeneous across different quantiles of the conditional distribution of dividends. In this study, we aim to contribute to the existing literature on dividend policy by incorporating location (quantile) asymmetries in the long-run target payout ratios and the dynamic dividend adjustment at the aggregate level. In order to construct a simple but flexible model for the dividend process that captures the stylised facts on management behavior, we apply the

asymmetries: the Fisher hypothesis is supported only in the upper quantiles. Furthermore, Wang (2012) studies inference on multiple structural breaks within quantile cointegrating regressions. Another related extension is to predictive quantile regression with highly persistent predictors (*e.g.*, Maynard *et al.*, 2011; Lee, 2014).

QARDL model to the dataset over the period from 1871Q3 to 2010Q2 extracted from Robert Shiller's web page (<http://www.econ.yale.edu/shiller>).

The full sample estimation results demonstrate that there is evidence of location asymmetries between lower and medium-to-higher quantiles of dividends. In particular, we find that the long-run payout ratio is higher and the dividend smoothing is stronger at higher quantiles than lower quantiles. Overall, our findings are consistent with the cross-sectional evidence by Leary and Michaely (2011) and the aggregate time series evidence in a global setting by Rangvid, Schmeling, and Schrimp (2012) that dividend smoothing is most common among large and mature firms with stable cash flows that are not financially constrained, face low levels of asymmetric information, and are readily susceptible to agency conflicts.⁴

We also allow for time-varying patterns of dividend policy by employing a rolling estimation technique with a window length of 320 quarters. A thorough examination of the time-varying QARDL estimation results provides a number of insightful findings on dividend policy in the US over the past century. First, dividend smoothing has become monotonically stronger over time. Similar monotonic downward trends have been observed for the impact coefficient with respect to contemporaneous changes in earnings. Both factors contribute to the extremely strong dividend smoothing reported in recent periods. Second, payout ratios have been monotonically decreasing over time, supporting the survey evidence by Brav *et al.* (2005) that the target payout ratio may no longer be the preeminent variable affecting payout decisions. Finally and more importantly, we find that the location asymmetries across different quantiles of the conditional distribution of dividends, which were clearly visible and pervasive in earlier periods, are less frequent in recent periods. These phenomena may indicate the establishment of financial deepening as a long-term process in the US, which serves to promote the stability of the whole financial system.

The paper is organized as follows. Section 2 introduces the QARDL model and derives the asymptotic distribution of both the short-run and the long-run estimators. Section 3 extends the QARDL model and its inference across multiple quantiles. Section 4 evaluates the finite-sample performance via Monte Carlo simulations. Section 5 presents the empirical application to an analysis of dividend smoothing and the long-run target payout ratio in the US. Section 6 presents concluding remarks and further research extensions. All of the proofs are relegated to the supplementary Technical Annex available at <http://web.yonsei.ac.kr/jinseocho/research.htm>.

Before proceeding, we discuss some notational details. A function is denoted using an empty argument.

⁴Denis and Osobov (2008) document that dividends are concentrated among the largest, most profitable payers in the US, Canada, UK, Germany, France, and Japan. In particular, they document that dividend payers account for more than 90% of the aggregate market capitalization except in the US and Canada, and the top 20% of dividend payers account for virtually all of the market capitalization of dividend payers. The concentration of dividends among the largest, most profitable firms is consistent with the life-cycle theory's central prediction that the distribution of free cash flow is the primary determinant of dividend policy.

Parameters without a subscript are used to indicate generic notations constituting the parameter space. Parameters attached with subscript or superscript “*” are those characterizing the data generating process. “ \Rightarrow ”, “ $\xrightarrow{\mathbb{P}}$ ”, and “ $[a]$ ” denote “weakly converges”, “converges in probability”, and “the smallest integer greater than a ”. We let $[a_{ij}]_{i=1,\dots,m,j=1,\dots,n}$ be an $m \times n$ matrix with an element a_{ij} and denote $\mathbf{1}_\ell$ as an $\ell \times 1$ vector of ones. $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard product of \mathbf{A} and \mathbf{B} . Other notations are standard.

2 Quantile Autoregressive Distributed Lag Model

The autoregressive distributed lag (ARDL) process can be extended to the quantile regression context. Note that the ARDL process is defined as $Y_t = \alpha_* + \sum_{j=1}^p \phi_{j*} Y_{t-j} + \sum_{j=0}^q \boldsymbol{\theta}_{j*}' \mathbf{X}_{t-j} + U_t$, where $\mathbf{X}_t \in \mathbb{R}^k$ is an integrated process of a stationary and ergodic process with population mean zero, U_t is the error term that is defined as $Y_t - \mathbb{E}[Y_t | \mathcal{F}_{t-1}]$ with \mathcal{F}_{t-1} being the smallest σ -field generated by $\{\mathbf{X}'_t, Y_{t-1}, \mathbf{X}'_{t-1}, \dots\}$, and p and q are lag orders. We further assume that the k variables in \mathbf{X}_t are not cointegrated among themselves. Following this framework, we let the τ -th quantile of Y_t conditional on \mathcal{F}_{t-1} be given as $\alpha_*(\tau) + \sum_{j=1}^p \phi_{j*}(\tau) Y_{t-j} + \sum_{j=0}^q \boldsymbol{\theta}_{j*}(\tau)' \mathbf{X}_{t-j}$ and denote this as $Q_{Y_t}(\tau | \mathcal{F}_{t-1})$. We now represent Y_t as

$$Y_t = \alpha_*(\tau) + \sum_{j=1}^p \phi_{j*}(\tau) Y_{t-j} + \sum_{j=0}^q \boldsymbol{\theta}_{j*}(\tau)' \mathbf{X}_{t-j} + U_t(\tau) \quad (1)$$

and call this the quantile autoregressive distributed lag (QARDL) process. Here, $U_t(\tau)$ is defined as $Y_t - Q_{Y_t}(\tau | \mathcal{F}_{t-1})$ as in Kim and White (2003).

To analyse the QARDL process we reformulate (1) as:

$$Y_t = \alpha_*(\tau) + \sum_{j=0}^{q-1} \mathbf{W}'_{t-j} \boldsymbol{\delta}_{j*}(\tau) + \mathbf{X}'_t \boldsymbol{\gamma}_*(\tau) + \sum_{j=1}^p \phi_{j*}(\tau) Y_{t-j} + U_t(\tau), \quad (2)$$

where $\boldsymbol{\gamma}_*(\tau) := \sum_{j=0}^q \boldsymbol{\theta}_{j*}(\tau)$, $\mathbf{W}_t := \Delta \mathbf{X}_t$, and $\boldsymbol{\delta}_{j*}(\tau) := -\sum_{i=j+1}^q \boldsymbol{\theta}_{i*}(\tau)$. All of the parameters in (2) measure the short-run dynamics, and the long-run relationship between Y_t and \mathbf{X}_t can be captured by reformulating (2) into the following long-run quantile process:

$$Y_t = \mu_*(\tau) + \mathbf{X}'_t \boldsymbol{\beta}_*(\tau) + \mathcal{R}_t(\tau) \quad \text{with} \quad (3)$$

$$\boldsymbol{\beta}_*(\tau) := \boldsymbol{\gamma}_*(\tau) \left(1 - \sum_{i=1}^p \phi_{i*}(\tau) \right)^{-1} \quad \text{and} \quad \mathcal{R}_t(\tau) := \sum_{j=0}^{\infty} \mathbf{W}'_{t-j} \boldsymbol{\xi}_{0,j*}(\tau) + \sum_{j=0}^{\infty} \rho_{j*}(\tau) U_{t-j}(\tau),$$

where we let $\mu_*(\tau) := \alpha_*(\tau) (1 - \sum_{i=1}^p \phi_{i*}(\tau))^{-1}$, $\boldsymbol{\xi}_{0,j*}(\tau) := \sum_{\ell=j+1}^{\infty} \boldsymbol{\pi}_{\ell*}(\tau)$, and $\{\rho_{0*}(\tau), \rho_{1*}(\tau), \dots\}$

and $\{\pi_{0*}(\tau), \pi_{1*}(\tau), \dots\}$ are such that $\sum_{j=0}^{\infty} \rho_{j*}(\tau)L^j \equiv (1 - \sum_{j=1}^p \phi_{j*}(\tau)L^j)^{-1}$ and

$$(1 - L)^{-1} \left(\frac{\sum_{j=0}^q \theta_{j*}(\tau)L^j}{1 - \sum_{j=1}^p \phi_{j*}(\tau)L^j} - \frac{\sum_{j=0}^q \theta_{j*}(\tau)}{1 - \sum_{j=1}^p \phi_{j*}(\tau)} \right) \equiv \sum_{j=0}^{\infty} \pi_{j*}(\tau)L^j,$$

respectively. Here, the static equation (3) is obtained by solving for Y_t from (2). The residual term $\mathcal{R}_t(\tau)$ represents the collection of serially correlated stationary variables irrelevant to the long-run relationship. We capture this by $\beta_*(\tau)$ and call it the *long-run parameter*. As is clear from its definition, $\beta_*(\tau)$ is defined as a function of $\gamma_*(\tau)$ and $\phi_*(\tau) := (\phi_{1*}(\tau), \dots, \phi_{p*}(\tau))'$, so that it can be consistently estimated by the plug-in principle and consistently estimating $\gamma_*(\tau)$ and $\phi_*(\tau)$.

Our main interests lie in developing the estimation theory for the long-run parameter $\beta_*(\tau)$. To this end, we reformulate the QARDL process in (2) as

$$Y_t = \mathbf{G}'_t \boldsymbol{\lambda}_*(\tau) + \tilde{\mathbf{Y}}'_t \boldsymbol{\phi}_*(\tau) + U_t(\tau) = \mathbf{Z}'_t \boldsymbol{\alpha}_*(\tau) + U_t(\tau), \quad (4)$$

where $\mathbf{Z}_t := (\mathbf{G}'_t, \tilde{\mathbf{Y}}'_t)' := (\mathbf{G}'_t, Y_{t-1}, \dots, Y_{t-p})' := (1, \mathbf{W}'_t, \dots, \mathbf{W}'_{t-q+1}, \mathbf{X}'_t, Y_{t-1}, \dots, Y_{t-p})'$, and $\boldsymbol{\alpha}_*(\tau) := [\boldsymbol{\lambda}_*(\tau)', \boldsymbol{\phi}_*(\tau)']' := [\alpha_*(\tau), \boldsymbol{\delta}_*(\tau)', \boldsymbol{\gamma}_*(\tau)', \boldsymbol{\phi}_*(\tau)']' := [\alpha_*(\tau), \boldsymbol{\delta}_{1*}(\tau)', \dots, \boldsymbol{\delta}_{(q-1)*}(\tau)', \boldsymbol{\gamma}_*(\tau)', \boldsymbol{\phi}_*(\tau)']'$. Here, each element of $\tilde{\mathbf{Y}}_t$ in (4) has the following specific form:

$$Y_{t-i} = \mu_*(\tau) + \mathbf{X}'_{t-i} \boldsymbol{\beta}_*(\tau) + \sum_{j=0}^{q-1} \mathbf{W}'_{t-j} \boldsymbol{\xi}_{i,j*}(\tau) + K_{t,i}(\tau), \quad i = 1, 2, \dots, p, \quad (5)$$

where we let $\boldsymbol{\xi}_{i,j*}(\tau) := -\boldsymbol{\beta}_*(\tau)$, if $i > j$; and $-\sum_{\ell=j-i}^{\infty} \boldsymbol{\pi}_{\ell*}(\tau)$, otherwise, and

$$K_{t,i}(\tau) := \begin{cases} -\sum_{j=q-i}^{\infty} \mathbf{W}'_{t-i-j} \boldsymbol{\xi}_{0,j*}(\tau) + \sum_{j=0}^{\infty} \rho_{j*}(\tau) U_{t-i-j}(\tau), & \text{if } i \leq q; \\ -\sum_{j=0}^{i-q-1} \mathbf{W}'_{t-q-j} \boldsymbol{\beta}_*(\tau) + \sum_{j=0}^{\infty} \mathbf{W}'_{t-i-j} \boldsymbol{\pi}_{j*}(\tau) + \sum_{j=0}^{\infty} \rho_{j*}(\tau) U_{t-i-j}(\tau), & \text{if } i > q. \end{cases}$$

(5) is the lagged version of (3). Pesaran and Shin (1998) provide a detailed derivation of (5). As (5) is iteratively used below, we rewrite it more compactly as $\tilde{\mathbf{Y}}_t = \boldsymbol{\Gamma}_*(\tau)' \mathbf{G}_t + \mathbf{K}_t(\tau)$ by letting

$$\boldsymbol{\Gamma}_*(\tau) := \begin{bmatrix} \mu_*(\tau) & \mu_*(\tau) & \cdots & \mu_*(\tau) \\ \boldsymbol{\xi}_{1,0*}(\tau) & \boldsymbol{\xi}_{2,0*}(\tau) & \cdots & \boldsymbol{\xi}_{p,0*}(\tau) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\xi}_{1,q-1*}(\tau) & \boldsymbol{\xi}_{2,q-1*}(\tau) & \cdots & \boldsymbol{\xi}_{p,q-1*}(\tau) \\ \boldsymbol{\beta}_*(\tau) & \boldsymbol{\beta}_*(\tau) & \cdots & \boldsymbol{\beta}_*(\tau) \end{bmatrix} \quad \text{and} \quad \mathbf{K}_t(\tau) := \begin{bmatrix} K_{t,1}(\tau) \\ K_{t,2}(\tau) \\ \vdots \\ K_{t,p-1}(\tau) \\ K_{t,p}(\tau) \end{bmatrix}.$$

We provide a couple of remarks about the quantile dependency of the QARDL parameters. First, we allow the short- and long-run parameters to be quantile-dependent, implying that the QARDL parameters can be affected by the innovation $U_t(\tau)$ received at each period and thus can be different across quantiles. Therefore, the (dynamic) conditioning variables not only shift the location, but also alter the scale and shape of the conditional distribution of Y_t . Furthermore, Xiao (2009) shows that the zero correlation between regressors and errors is a key condition for consistently estimating the quantile-dependent cointegrating vector. Notice that this is the main reason why Xiao allows the quantile-dependent cointegrating parameters only under the dynamic OLS approach using leads and lags, not under the semiparametric fully-modified framework. In the ARDL context, this condition is equivalent to $\mathbb{E}[\Delta \mathbf{X}_t U_t] = 0$ and $\mathbb{E}[Y_{t-j} U_t] = 0$ ($j = 1, \dots, p$) that typically hold if sufficiently large lag orders are given for p and q . Second, it is straightforward to rewrite the QARDL process, (1) in the following error correction model (ECM) form:

$$\Delta Y_t = \alpha_*(\tau) + \zeta_*(\tau) (Y_{t-1} - \beta_*(\tau)' \mathbf{X}_{t-1}) + \sum_{j=1}^{p-1} \phi_j^*(\tau) \Delta Y_{t-j} + \sum_{j=0}^{q-1} \theta_j^*(\tau)' \Delta \mathbf{X}_{t-j} + U_t(\tau), \quad (6)$$

where $\zeta_*(\tau) := \sum_{i=1}^p \phi_{i*}(\tau) - 1$, $\theta_0^*(\tau) = \theta_{0*}(\tau)$, and for $j = 1, \dots, p-1$, $\phi_j^*(\tau) := -\sum_{h=j+1}^p \phi_{h*}(\tau)$ and $\theta_j^*(\tau) := -\sum_{h=j+1}^p \theta_{h*}(\tau)$. We refer to (6) as the QARDL-ECM representation. The special case of this is the quantile-invariant homogeneous cointegration with $\beta_*(\tau) = \beta_*$ for all τ 's. Even for this simple case, we may still be interested in testing whether the speed of adjustment $\zeta_*(\tau)$ is quantile dependent.

We now define our estimators. First, the reformulation in (4) can be used to estimate the unknown short-run parameters by applying the standard quantile regression approach, *e.g.*, Koenker and Hallock (2001). For a given $\tau \in (0, 1)$, we obtain the QARDL estimator as $\tilde{\alpha}_n(\tau) := \arg \min_{\alpha(\tau)} \sum_{t=1}^n \varrho_\tau \{Y_t - \mathbf{Z}_t' \alpha(\tau)\}$, where $\varrho_\tau(u) := u\psi_\tau(u)$ is the so-called *check function* with $\psi_\tau(u) := \tau - I(u \leq 0)$. Conformably with $\alpha_*(\tau)$, we partition $\tilde{\alpha}_n(\tau)$ into $\tilde{\alpha}_n(\tau) := (\tilde{\lambda}_n(\tau)', \tilde{\phi}_n(\tau)')' := (\tilde{\alpha}_n(\tau), \tilde{\delta}_n(\tau)', \tilde{\gamma}_n(\tau)', \tilde{\phi}_n(\tau)')'$ for future references. Next, we estimate the long-run parameter using the plug-in principle: $\tilde{\beta}_n(\tau) := \tilde{\gamma}_n(\tau)(1 - \sum_{j=1}^p \tilde{\phi}_{j,n}(\tau))^{-1}$. It is clear from this definition that the behavior of $\tilde{\beta}_n(\tau)$ governing the quantile long-run equilibrium is intrinsically related to the short-run estimators $\tilde{\phi}_{j,n}(\tau)$ and $\tilde{\gamma}_n(\tau)$.

Recently, Xiao (2009) advances a quantile cointegration approach in a static regression context and develops both the semiparametric fully modified estimator and the parametrically augmented estimator using leads and lags of the first-differenced regressors. These correspond to the quantile extensions of the estimators proposed by Phillips and Hansen (1990) and by Saikkonen (1991), respectively. These estimators are asymptotically shown to follow the mixture normal distributions even in the presence of endogenous regressors and/or serially correlated residuals. We also note that our proposed QARDL-ECM approach is

different from Xiao's (2009) approach, which is typically built upon Engle and Granger's (1987) static cointegrating regression. On the other hand, the QARDL estimator simultaneously addresses both the long-run relationship and the associated short-run dynamics across a wide range of quantiles.

We impose the following assumptions to examine the regular properties of the QARDL estimator.

Assumption 1. (i) $\{U_t\}$ is an IID sequence with finite variance such that it has a continuous probability density function (PDF) $f(\cdot)$ and cumulative density function (CDF) $F(\cdot)$ with $f(\cdot) > 0$ and $f_\tau := f[F^{-1}(\tau)] < \infty$; (ii) \mathbf{X}_t is a $k \times 1$ vector of integrated regressors such that $\mathbf{W}_t := \Delta \mathbf{X}_t$ is a general linear multivariate stationary and ergodic process with $\mathbb{E}[W_{tj}] = 0$ and $\mathbb{E}[|W_{tj}|^2] < \infty$ for $j = 1, 2, \dots, k$;

(iii) For each $\tau \in (0, 1)$ and $t, s = 1, 2, \dots$, $U_t(\tau)$ and \mathbf{W}_s are independent;

(iv) The cointegration order of \mathbf{X}_t is zero;

(v) For each $\tau \in (0, 1)$, the roots of $(1 - \sum_{j=1}^p \phi_{j*}(\tau)L^j)$ lie outside the unit circle, and for all $i = 1, \dots, k$, $\sum_{j=0}^{\infty} |\xi_{0,j,i*}(\tau)| < \infty$ and $\sum_{j=0}^{\infty} |\pi_{j,i*}(\tau)| < \infty$, where $\xi_{0,j,i*}(\tau)$ and $\pi_{j,i*}(\tau)$ are the i -th elements of $\boldsymbol{\xi}_{0,j*}(\tau)$ and $\boldsymbol{\pi}_{j*}(\tau)$, respectively; and

(vi) For each $r, \tau \in (0, 1)$, $\mathbf{B}_n(\cdot, \tau) \Rightarrow \mathcal{B}(\cdot, \tau)$, where $\mathbf{B}_n(r, \tau) := \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} [\bar{\mathbf{W}}'_t, \bar{\mathbf{K}}_t(\tau)', \psi_\tau[U_t(\tau)], \psi_\tau[U_t(\tau)]\bar{\mathbf{W}}'_t, \psi_\tau[U_t(\tau)]\mathbf{K}_t(\tau)']'$, $\bar{\mathbf{K}}_t(\tau) := \mathbf{K}_t(\tau) - \mathbb{E}[\mathbf{K}_t(\tau)]$, $\bar{\mathbf{W}}_t := [\mathbf{W}'_t, \mathbf{W}'_{t-1}, \dots, \mathbf{W}'_{t-q+1}]'$, and $\mathcal{B}(\cdot, \tau) := [\mathcal{B}_{\mathbf{W}}(\cdot)', \mathcal{B}_{\mathbf{K}}(\cdot, \tau)', \mathcal{B}_\psi(\cdot, \tau), \mathcal{B}_{\psi \cdot \mathbf{W}}(\cdot, \tau)', \mathcal{B}_{\psi \cdot \mathbf{K}}(\cdot, \tau)']'$ is a multivariate Gaussian process such that $\mathbb{E}[\mathcal{B}(\cdot, \tau)] = \mathbf{0}$, and for each $r, \tilde{r} \in (0, 1)$, $\mathbb{E}[\mathcal{B}(r, \tau) \mathcal{B}(\tilde{r}, \tau)'] = \boldsymbol{\Omega}(r, \tilde{r}, \tau)$ with $\boldsymbol{\Omega}(r, \tilde{r}, \tau) := \lim_{n \rightarrow \infty} \mathbb{E}[\mathbf{B}_n(r, \tau) \mathbf{B}_n(\tilde{r}, \tau)']$ being positive definite. \square

A number of remarks are in order. First, the infinite density case is eliminated by Assumption 1(i).⁵ Second, $\{U_t\}$ is assumed to be an IID process for analytical convenience. Allowing $\{U_t\}$ to be a martingale difference array (MDA) does not alter the main results, although more tedious derivations are required. In addition, the IID condition of Assumption 1(i) implies that $f_{U_t|\mathbf{Z}_t}(\cdot|\mathbf{Z}_t) = f(\cdot)$. Third, Assumption 1(v) is imposed for the existence of a stable long-run relationship between Y_t and \mathbf{X}_t . Fourth, the structure of $\boldsymbol{\Omega}(r, \tilde{r}, \tau)$ in Assumption 1(vi) is obtained as $\lim_{t \rightarrow \infty} \mathbb{E}[\mathbf{B}_n(r, \tau) \mathbf{B}_n(\tilde{r}, \tau)']$. In particular, $\mathcal{B}_{\mathbf{W}}(\cdot)$, $\mathcal{B}_\psi(\cdot, \tau)$ and $\mathcal{B}_{\psi \cdot \mathbf{W}}(\cdot, \tau)$ are independent according to Assumption 1(iii) and the fact that $\mathbb{E}[\psi_\tau[U_t(\tau)]] = 0$, $\mathbb{E}[\bar{\mathbf{W}}_s] = \mathbf{0}$, $\mathbb{E}[\bar{\mathbf{W}}_s \psi_\tau[U_t(\tau)]] = \mathbf{0}$, $\mathbb{E}[\bar{\mathbf{W}}_s \bar{\mathbf{W}}'_s \psi_\tau[U_t(\tau)]] = \mathbf{0}$, and $\mathbb{E}[\bar{\mathbf{W}}_s \psi_\tau[U_t(\tau)]^2] = \mathbf{0}$. This renders the covariances of $\mathcal{B}_{\mathbf{W}}(\cdot)$, $\mathcal{B}_\psi(\cdot, \tau)$ and $\mathcal{B}_{\psi \cdot \mathbf{W}}(\cdot, \tau)$ to be zero. On the other hand, $\mathbf{K}_s(\tau)$ is not necessarily independent of $U_t(\tau)$, although $\mathbf{K}_s(\tau)$ is a vector of stationary and ergodic processes based on the definition of $K_{t,i}(\tau)$ and Assumptions 1(ii, iii). This aspect makes the covariance between $\mathcal{B}_{\mathbf{K}}(\cdot, \tau)'$ and $\mathcal{B}_{\psi \cdot \mathbf{K}}(\cdot, \tau)'$ indeterministic. We impose the positive-definite matrix condition to $\boldsymbol{\Omega}(r, \tilde{r}, \tau)$ as its minimal condition.

⁵Cho, Han, and Phillips (2010) and Han, Cho, and Phillips (2011) examine the asymptotic behavior of the LAD estimator when the density function has infinite value at the median using the methodology of Knight (1998).

This is a technical condition that usually holds for most empirical data. Fifth, by the definition of \mathbf{X}_t , it is straightforward to derive that $n^{-1} \sum_{t=1}^n \psi_\tau[U_t(\tau)] \mathbf{X}_t \Rightarrow \int_0^1 \bar{\mathbf{B}}_W(r) d\mathcal{B}_\psi(r, \tau)$ and $n^{-2} \sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t' \Rightarrow \int_0^1 \bar{\mathbf{B}}_W(r) \bar{\mathbf{B}}_W(r)' dr$, where $\bar{\mathbf{B}}_W(\cdot)$ is the first k elements of $\mathbf{B}_W(\cdot)$. Since $\psi_\tau[U_t(\tau)]$ is independent of \mathbf{W}_t , the one-sided long-run covariance is equal to zero. Sixth, the independence condition in Assumption 1(iii) is not as strong as it appears. In particular, the empirical application in Section 5 illustrates how one can control for endogeneity of regressors by projecting the regression errors on $\Delta \mathbf{X}_t$. See also Pesaran and Shin (1998) and Pesaran *et al.* (2001). We, therefore, stick to this condition without loss of generality of our analysis.⁶ Seventh, we impose the positive-definite condition to $\boldsymbol{\Omega}(r, \tilde{r}, \tau)$ to prevent identical or constant sample paths in $\mathcal{B}(\cdot, \tau)$. Eighth, the high level condition in Assumption 1(vi) can be attained in a couple of ways. Given the other conditions, if $\{(\bar{\mathbf{W}}_t', \mathbf{K}_t(\tau)')'\}$ is an adapted mixingale of size -1 with finite global variance, Theorem 3 in Scott (1973) provides the desired result. Alternatively, Assumption 1(vi) follows if the summability condition on Assumption 1(v) is further strengthened and their fourth moment is finite. See for example Phillips and Solo (1992) for this derivation. As one of these two assumptions does not necessarily imply the other, we opt to impose the high level assumption. Finally, Assumptions 1(i and iii) imply that $E[\psi_\tau(U_t(\tau)) | Y_{t-1}, Y_{t-2}, \dots, \mathbf{X}_t, \mathbf{X}_{t-1}, \dots] = 0$, so that the τ -th quantile of the conditional distribution of $U_t(\tau)$ is zero. Hence, the coefficients of the QARDL model in (1) are identified.

We provide the asymptotic distributions of the short-run estimators $\tilde{\phi}_{j,n}(\tau)$ and $\tilde{\gamma}_n(\tau)$ below.

Theorem 1 (Short-Run Estimators). *Under Assumption 1,*

- (i) For each $\tau \in (0, 1)$, $\sqrt{n}(\tilde{\phi}_n(\tau) - \phi_*(\tau)) \Rightarrow N[\mathbf{0}, \boldsymbol{\Pi}(\tau)]$, where $\boldsymbol{\Pi}(\tau) := \tau(1 - \tau)f_\tau^{-2} \mathbb{E}[\tilde{\mathbf{H}}_t(\tau) \tilde{\mathbf{H}}_t(\tau)']^{-1}$, $\tilde{\mathbf{H}}_t(\tau) := \mathbf{K}_t(\tau) - \mathbb{E}[\mathbf{K}_t(\tau) \tilde{\mathbf{W}}_t'] \mathbb{E}[\tilde{\mathbf{W}}_t \tilde{\mathbf{W}}_t']^{-1} \tilde{\mathbf{W}}_t$, and $\tilde{\mathbf{W}}_t := [1, \bar{\mathbf{W}}_t']'$; and
- (ii) For each $\tau \in (0, 1)$, $\sqrt{n}(\tilde{\gamma}_n(\tau) - \gamma_*(\tau)) \Rightarrow N[\mathbf{0}, \boldsymbol{\beta}_*(\tau) \boldsymbol{\nu}_p' \boldsymbol{\Pi}(\tau) \boldsymbol{\nu}_p \boldsymbol{\beta}_*(\tau)']$. □

Therefore, $(\tilde{\phi}_n(\tau)', \tilde{\gamma}_n(\tau)')'$ is consistent for $(\phi_*(\tau)', \gamma_*(\tau)')'$, and its convergence rate is $n^{1/2}$. We outline the key steps of the proof. Using Assumption 1, we show that

$$\sqrt{n}(\tilde{\phi}_n(\tau) - \phi_*(\tau)) = f_\tau^{-1} \mathbb{E}[\tilde{\mathbf{H}}_t(\tau) \tilde{\mathbf{H}}_t(\tau)']^{-1} \left(n^{-1/2} \tilde{\mathbf{H}}(\tau)' \boldsymbol{\Psi}_\tau(\mathbf{U}) \right) + o_{\mathbb{P}}(1), \quad (7)$$

where $\tilde{\mathbf{H}}(\tau) := [\tilde{\mathbf{H}}_1(\tau), \dots, \tilde{\mathbf{H}}_n(\tau)]'$, and $\boldsymbol{\Psi}_\tau(\mathbf{U}) := [\psi_\tau[U_1(\tau)], \dots, \psi_\tau[U_n(\tau)]]'$. It is also straightforward to show that

$$n^{-1/2} \tilde{\mathbf{H}}(\tau)' \boldsymbol{\Psi}_\tau(\mathbf{U}) \Rightarrow N[\mathbf{0}, \tau(1 - \tau) \mathbb{E}[\tilde{\mathbf{H}}_t(\tau) \tilde{\mathbf{H}}_t(\tau)']] \quad (8)$$

by applying the central limit theorem (CLT) for MDA. Theorem 1(i) follows by combining (7) and (8).

⁶We also note that the independence assumption can be relaxed into the zero correlation condition, as in OLS regression, in which case the variance-covariance matrix will be expressed in the usual sandwich form.

Pesaran and Shin (1998) derive the asymptotic covariance matrix in their context by implicitly imposing that $\mathbb{E}[\mathbf{K}_t \widetilde{\mathbf{W}}_t']$ is zero. However, this is too strong unless $\{\mathbf{W}_t\}$ is an IID process. Hence, we relax this restriction in what follows. Second, we can prove Theorem 1(ii) by first showing that

$$\sqrt{n} \left\{ (\tilde{\gamma}_n(\tau) - \gamma_*(\tau)) + \boldsymbol{\beta}_*(\tau) \boldsymbol{\nu}'_p(\tilde{\phi}_n(\tau) - \phi_*(\tau)) \right\} = o_{\mathbb{P}}(1), \quad (9)$$

and then using the asymptotic distribution of $\sqrt{n}(\tilde{\phi}_n(\tau) - \phi_*(\tau))$ given in Theorem 1(i). It is clear from (9) that the asymptotic distribution of $\tilde{\gamma}_n(\tau)$ depends upon that of $\tilde{\phi}_n(\tau)$.

We examine the asymptotic behaviors of the long-run parameter estimator in the following theorem.

Theorem 2 (Long-Run Estimator). *Under Assumption 1, and for each $\tau \in (0, 1)$,*

(i) $n(\tilde{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_*(\tau)) \Rightarrow \{f_\tau(1 - \sum_{j=1}^p \phi_{j*}(\tau))\}^{-1} [\int_0^1 \tilde{\mathbf{B}}_W(r) \tilde{\mathbf{B}}_W(r)' dr]^{-1} \int_0^1 \tilde{\mathbf{B}}_W(r) d\mathcal{B}_\psi(r, \tau)$, where $\tilde{\mathbf{B}}_W(r) := \bar{\mathbf{B}}_W(r) - \int_0^1 \bar{\mathbf{B}}_W(r) dr$; and

(ii) $n\mathbf{M}^{1/2}(\tilde{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_*(\tau)) \Rightarrow N[\mathbf{0}, \tau(1 - \tau)\{f_\tau(1 - \sum_{j=1}^p \phi_{j*}(\tau))\}^{-2} \mathbf{I}_k]$, where $\mathbf{M} := n^{-2} \mathbf{X}'[\mathbf{I} - \widetilde{\mathbf{W}}(\widetilde{\mathbf{W}}' \widetilde{\mathbf{W}})^{-1} \widetilde{\mathbf{W}}'] \mathbf{X}$ and $\widetilde{\mathbf{W}} := [\widetilde{\mathbf{W}}_1, \widetilde{\mathbf{W}}_2, \dots, \widetilde{\mathbf{W}}_n]'$. \square

Theorem 2(i) implies that $\tilde{\boldsymbol{\beta}}_n(\tau)$ is consistent for $\boldsymbol{\beta}_*(\tau)$ at the convergence rate of n . Notice that the normality result of Theorem 2(ii) is important for testing the hypotheses as developed below, and it is mainly obtained by the independence condition between $\widetilde{\mathbf{W}}$ and $\boldsymbol{\Psi}_\tau(\mathbf{U})$ by Assumption 1(iii). More specifically, the key steps of our proof are described as follows. First, we note the following identity:

$$\left(\tilde{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_*(\tau) \right) \equiv \left[1 - \sum_{j=1}^p \tilde{\phi}_{j,n}(\tau) \right]^{-1} \left[(\tilde{\gamma}_n(\tau) - \gamma_*(\tau)) + \boldsymbol{\beta}_*(\tau) \sum_{j=1}^p (\tilde{\phi}_{j,n}(\tau) - \phi_{j*}(\tau)) \right]. \quad (10)$$

Using this identity, it follows that $n\{(\tilde{\gamma}_n(\tau) - \gamma_*(\tau)) + \boldsymbol{\beta}_*(\tau) \sum_{j=1}^p (\tilde{\phi}_{j,n}(\tau) - \phi_{j*}(\tau))\} = f_\tau^{-1} \mathbf{M}^{-1} \{n^{-1} \mathbf{X}'[\mathbf{I} - \widetilde{\mathbf{W}}(\widetilde{\mathbf{W}}' \widetilde{\mathbf{W}})^{-1} \widetilde{\mathbf{W}}'] \boldsymbol{\Psi}_\tau(\mathbf{U})\} + o_{\mathbb{P}}(1)$. This holds by noting the fact that $n\{(\tilde{\gamma}_n(\tau) - \gamma_*(\tau)) + \boldsymbol{\beta}_*(\tau) \sum_{j=1}^p (\tilde{\phi}_{j,n}(\tau) - \phi_{j*}(\tau))\}$ is equivalent to the last k elements of $\mathbf{D}_G \{(\tilde{\boldsymbol{\lambda}}_n(\tau) - \boldsymbol{\lambda}_*(\tau)) + \boldsymbol{\Gamma}_*(\tau)(\tilde{\phi}_n(\tau) - \phi_*(\tau))\}$ in probability, where $\mathbf{D}_G := \text{diag}([\sqrt{n} \boldsymbol{\nu}'_{1+qk}, n \boldsymbol{\nu}'_k]')$. Furthermore, the latter is equal to $f_\tau^{-1} \mathbf{M}^{-1} \{n^{-1} \mathbf{X}'[\mathbf{I} - \widetilde{\mathbf{W}}(\widetilde{\mathbf{W}}' \widetilde{\mathbf{W}})^{-1} \widetilde{\mathbf{W}}'] \boldsymbol{\Psi}_\tau(\mathbf{U})\}$ in probability. As the right-hand side (RHS) of this equation is bounded in probability, we can apply Theorem 1(i) and obtain that $n(\tilde{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_*(\tau)) = [1 - \sum_{j=1}^p \tilde{\phi}_{j,n}(\tau)]^{-1} (f_\tau \mathbf{M})^{-1} \{n^{-1} \mathbf{X}'[\mathbf{I} - \widetilde{\mathbf{W}}(\widetilde{\mathbf{W}}' \widetilde{\mathbf{W}})^{-1} \widetilde{\mathbf{W}}'] \boldsymbol{\Psi}_\tau(\mathbf{U})\} + o_{\mathbb{P}}(1)$. As a result, the asymptotic distribution in Theorem 2(i) is established. Next, Theorem 2(ii) holds as a corollary of Theorem 2(i). By Assumption 1(iii), $\widetilde{\mathbf{W}}$ and $\boldsymbol{\Psi}_\tau(\mathbf{U})$ are independent, and thus the weak limit in Theorem 2(i) can be viewed as a mixture of normal random variables. Theorem 2(ii) captures this implication, and its standardized version asymptotically

follows a multivariate normal distribution.

The mixture normal distribution property for the long-run parameter estimator implies that the Wald statistic asymptotically follows a chi-squared distribution under a suitable null hypothesis for $\beta_*(\tau)$. To investigate this property, we consider the following null and alternative hypotheses:

$$\mathbb{H}_0^{(1)} : \mathbf{R}\beta_*(\tau) = \mathbf{r} \text{ versus } \mathbb{H}_1^{(1)} : \mathbf{R}\beta_*(\tau) \neq \mathbf{r},$$

where \mathbf{R} is an $r \times k$ matrix and \mathbf{r} is an $r \times 1$ vector. The Wald statistic is defined as

$$\mathcal{W}_n(\beta) := \frac{n^2 \widehat{f}_\tau^2}{\tau(1-\tau)} \left(1 - \sum_{j=1}^p \widetilde{\phi}_{j,n}(\tau) \right)^2 \left(\mathbf{R}\widetilde{\beta}_n(\tau) - \mathbf{r} \right)' \left(\mathbf{R}\mathbf{M}^{-1}\mathbf{R}' \right)^{-1} \left(\mathbf{R}\widetilde{\beta}_n(\tau) - \mathbf{r} \right),$$

where \widehat{f}_τ is a consistent estimator of f_τ . Using the definition of $\widetilde{\beta}_n(\tau)$, the Wald statistic can be rewritten as

$$\frac{n^2 \widehat{f}_\tau^2}{\tau(1-\tau)} \left[\mathbf{R}\widetilde{\gamma}_n(\tau) - \left(1 - \sum_{j=1}^p \widetilde{\phi}_{j,n}(\tau) \right) \mathbf{r} \right]' \left(\mathbf{R}\mathbf{M}^{-1}\mathbf{R}' \right)^{-1} \left[\mathbf{R}\widetilde{\gamma}_n(\tau) - \left(1 - \sum_{j=1}^p \widetilde{\phi}_{j,n}(\tau) \right) \mathbf{r} \right].$$

Note that this Wald test statistic is a generalized version of the t -test statistic given by

$$\mathcal{T}_{m,n} := n \left(1 - \sum_{j=1}^p \widetilde{\phi}_{j,n}(\tau) \right) \frac{\widehat{f}_\tau (\mathbf{R}_m \widetilde{\beta}_n(\tau) - r_m)}{\{\tau(1-\tau) \mathbf{R}_m \mathbf{M}^{-1} \mathbf{R}_m'\}^{1/2}},$$

where $\mathbf{R}_m := [\mathbf{0}'_{(m-1) \times 1}, 1, \mathbf{0}'_{(k-m) \times 1}]$ and r_m is a constant. It is easily seen that $\mathcal{T}_{m,n}^2$ equals the Wald statistic that tests $\mathbb{H}_0^{(0)} : \beta_{m*}(\tau) = r_m$ against $\mathbb{H}_1^{(0)} : \beta_{m*}(\tau) \neq r_m$, where $\beta_{m*}(\tau)$ is an m -th element of $\beta_*(\tau)$.

The following corollary provides the asymptotic behaviors of $\mathcal{W}(\beta)$ and $\mathcal{T}_{m,n}$ statistics.

Corollary 1. *Suppose that Assumption 1 holds.*

(i) *For each $m = 1, \dots, k$,*

(i.a) $\mathcal{T}_{m,n} \Rightarrow N(0, 1)$ *under* $\mathbb{H}_0^{(0)}$;

(i.b) $\mathbb{P}(|\mathcal{T}_{m,n}| \geq c_n) \rightarrow 1$ *for any sequence* $\{c_n\}$ *s.t.* $c_n = o(n)$ *under* $\mathbb{H}_1^{(0)}$; *and*

(ii) *If* $\text{rank}(\mathbf{R}) = r$ *and* $\widehat{f}_\tau \xrightarrow{\mathbb{P}} f_\tau$,

(ii.a) $\mathcal{W}_n(\beta) \Rightarrow \chi_r^2$ *under* $\mathbb{H}_0^{(1)}$; *and*

(ii.b) $\mathbb{P}(\mathcal{W}_n(\beta) \geq c'_n) \rightarrow 1$ *for any sequence* $\{c'_n\}$ *s.t.* $c'_n = o(n^2)$ *under* $\mathbb{H}_1^{(1)}$. □

Corollary 1 is a straightforward consequence of Theorem 2(ii). In particular, $\sum_{j=1}^p \widetilde{\phi}_{j,n}(\tau)$ consistently

estimates $\sum_{j=1}^p \phi_{j*}(\tau)$ by Theorem 1 under Assumption 1. Thus, employing $\tilde{\phi}_n(\tau)$ to define the test statistics does not alter their asymptotic null distributions. We can also see that both t -test and Wald tests have nontrivial powers against the local alternatives converging to the null at the rate of n^{-1} by associating Corollaries 1(i.b) and (ii.b) along with the convergence rate of each test under the null hypothesis.

We construct statistics for testing linear restrictions on the short-run parameters and consider the following null and alternative hypotheses:

$$\mathbb{H}_0^{(2)} : \mathbf{Q}\phi_*(\tau) = \mathbf{q} \text{ versus } \mathbb{H}_1^{(2)} : \mathbf{Q}\phi_*(\tau) \neq \mathbf{q}, \quad \text{and} \quad (11)$$

$$\mathbb{H}_0^{(3)} : \mathbf{R}\gamma_*(\tau) = \mathbf{r} \text{ versus } \mathbb{H}_1^{(3)} : \mathbf{R}\gamma_*(\tau) \neq \mathbf{r}, \quad (12)$$

where \mathbf{Q} is an $r \times p$ matrix, \mathbf{q} is an $r \times 1$ vector, and $\text{rank}(\mathbf{R}) = 1$ that is needed to prevent the Wald test from being degenerate. The asymptotic null distributions of the Wald statistics that test the hypotheses in (11) and (12) are regular by Theorem 1. We formally define the following Wald test statistic:

$$\mathcal{W}_n(\phi) := n \left(\mathbf{Q}\tilde{\phi}_n(\tau) - \mathbf{q} \right)' \left(\mathbf{Q}\hat{\Pi}_n(\tau)\mathbf{Q}' \right)^{-1} \left(\mathbf{Q}\tilde{\phi}_n(\tau) - \mathbf{q} \right).$$

Here, $\hat{\Pi}_n(\tau)$ is any consistent estimator for $\Pi(\tau) := \tau(1-\tau)f_\tau^{-2}\mathbb{E}[\tilde{\mathbf{H}}_t\tilde{\mathbf{H}}_t']^{-1}$. By definition, $\Pi(\tau)$ can be estimated in a number of ways. As an example, for a given consistent estimator \hat{f}_τ for f_τ , we let:

$$\hat{\Pi}_n(\tau) := \hat{f}_\tau^{-1}\tau(1-\tau) \left\{ \left(n^{-1}\hat{\mathbf{K}}(\tau)'\hat{\mathbf{K}}(\tau) \right) - \left(n^{-1}\hat{\mathbf{K}}(\tau)'\tilde{\mathbf{W}} \right) \left(n^{-1}\tilde{\mathbf{W}}'\tilde{\mathbf{W}} \right)^{-1} \left(n^{-1}\tilde{\mathbf{W}}'\hat{\mathbf{K}}(\tau) \right) \right\}, \quad (13)$$

where $\hat{\mathbf{K}}(\tau) := [\hat{\mathbf{K}}_1, \hat{\mathbf{K}}_2, \dots, \hat{\mathbf{K}}_n]'$, $\hat{\mathbf{K}}_t(\tau) := [\hat{K}_{t,1}(\tau), \hat{K}_{t,2}(\tau), \dots, \hat{K}_{t,p}(\tau)]'$, and for each i , $\hat{K}_{t,i}(\tau)$ is the quantile regression error obtained from regressing $\tilde{\mathbf{W}}_t$ against $Y_{t-i} - \mathbf{X}'_t\tilde{\beta}_n(\tau)$. By (5), for each t , $\hat{\mathbf{K}}_t(\tau)$ is a consistent quantile regression error for $\mathbf{K}_t(\tau)$, and this leads to $\hat{\Pi}_n(\tau) \xrightarrow{\mathbb{P}} \Pi(\tau)$. In addition, many other estimators can be similarly defined. The following corollary provides the asymptotic behaviors of the Wald statistic $\mathcal{W}(\phi)$.

Corollary 2. *Suppose that Assumption 1 holds and $\hat{\Pi}_n(\tau) \xrightarrow{\mathbb{P}} \Pi(\tau)$. If $\text{rank}(\mathbf{Q}) = r$,*

(i) $\mathcal{W}_n(\phi) \Rightarrow \chi_r^2$ under $\mathbb{H}_0^{(2)}$; and

(ii) $\mathbb{P}(\mathcal{W}_n(\phi) \geq c_n'') \rightarrow 1$ for any sequence $\{c_n''\}$ s.t. $c_n'' = o(n)$ under $\mathbb{H}_1^{(2)}$.

Under the conditions in Corollary 2, it is straightforward to prove Corollary 2 using Theorem 1. We, therefore, omit the proof. Also, the Wald test in Corollary 2 has a nontrivial power against the local alternatives converging to the null at the rate of $n^{-1/2}$ because the convergence rate of $\tilde{\phi}_n(\tau)$ is $n^{-1/2}$. Furthermore, as

the speed of adjustment or error correction coefficient is defined by $\zeta_*(\tau) := \sum_{i=1}^p \phi_{i*}(\tau) - 1$ in (6), we can test the null hypothesis that this coefficient is nil by letting $\mathbf{Q} = [1, \dots, 1]$ and $\mathbf{q} = 1 + \zeta_0$.

We now examine the asymptotic behaviors of the Wald statistic that tests $\mathbb{H}_0^{(3)}$ versus $\mathbb{H}_1^{(3)}$ under the null and alternative hypotheses in (12). The asymptotic distribution of $\sqrt{n}(\tilde{\gamma}_n(\tau) - \gamma_*(\tau))$ is degenerate because its asymptotic covariance $\beta_*(\tau)\boldsymbol{\nu}'_p\boldsymbol{\Pi}(\tau)\boldsymbol{\nu}_p\beta'_*(\tau)$ has a rank equal to unity by Theorem 1(ii). This implies that, if \mathbf{R} has a rank greater than unity, the corresponding Wald test has a degenerate asymptotic null distribution. That is why we let $\text{rank}(\mathbf{R}) = 1$. The Wald test statistic is defined as

$$\mathcal{W}_n(\gamma) := n(\mathbf{R}\tilde{\gamma}_n(\tau) - \mathbf{r})' \left(\mathbf{R}\tilde{\beta}_n(\tau)\boldsymbol{\nu}'_p\hat{\boldsymbol{\Pi}}_n(\tau)\boldsymbol{\nu}_p\tilde{\beta}'_n(\tau)\mathbf{R}' \right)^{-1} (\mathbf{R}\tilde{\gamma}_n(\tau) - \mathbf{r}).$$

The asymptotic covariance matrix is estimated by estimating $\beta_*(\tau)$ and $\boldsymbol{\Pi}(\tau)$ through $\tilde{\beta}_n(\tau)$ and $\hat{\boldsymbol{\Pi}}_n(\tau)$, respectively. The regular asymptotic behaviors of this Wald test statistic can be summarised as follows:

Corollary 3. *Suppose that Assumption 1 holds and $\hat{\boldsymbol{\Pi}}_n(\tau) \xrightarrow{\mathbb{P}} \boldsymbol{\Pi}(\tau)$.*

- (i) $\mathcal{W}_n(\gamma) \Rightarrow \chi_1^2$ under $\mathbb{H}_0^{(3)}$; and
- (ii) $\mathbb{P}(\mathcal{W}_n(\gamma) \geq c_n''') \rightarrow 1$ for any sequence $\{c_n'''\}$ s.t. $c_n''' = o(n)$ under $\mathbb{H}_1^{(3)}$.

Corollary 3 is trivially implied by Theorem 1(ii).

There are additional remarks relevant to Corollary 3. As pointed out above, it is difficult to apply $\mathcal{W}_n(\gamma)$ if the rank of \mathbf{R} is greater than unity. Nevertheless, this difficulty does not arise if \mathbf{R} and \mathbf{r} are designed to test the ratios of the elements in $\gamma_*(\tau)$. Note that $\gamma_*(\tau) \equiv \{1 - \sum_{j=1}^p \phi_{j*}(\tau)\}\beta_*(\tau)$ by definition, so that a ratio of the elements in $\gamma_*(\tau)$ is virtually the ratio of the elements in $\beta_*(\tau)$. This also implies that the same hypothesis can be tested using $\mathcal{W}_n(\beta)$. Similarly to $\mathcal{W}_n(\phi)$, $\mathcal{W}_n(\gamma)$ has a nontrivial power against the local alternatives that converge to the null at the rate, $n^{-1/2}$.

3 Inference with Multiple Quantiles

In the quantile regression, we typically estimate the model using multiple quantile indices. In this case, one may wish to test the null that the short- or the long-run parameters at a low quantile ($\tau = 0.1$) are the same as those at an upper quantile ($\tau = 0.9$). Any evidence of disparity can be regarded as an indication of an asymmetric behavior associated with the distributional location of the dependent variable.

We develop a systematic inference procedure for testing the hypotheses that are constructed by multiple quantile indices. Consider an s number of quantile indices, say $\tau_1 < \dots < \tau_s$, and let their corresponding multi-quantile short-run and long-run parameters be $[\boldsymbol{\Phi}_*(\boldsymbol{\tau})', \boldsymbol{\Gamma}_*(\boldsymbol{\tau})']' := [\phi_*(\tau_1)', \dots, \phi_*(\tau_s)', \gamma_*$

$(\tau_1)', \dots, \gamma_*(\tau_s)']'$ and $\mathbf{B}_*(\boldsymbol{\tau}) := [\boldsymbol{\beta}_*(\tau_1)', \dots, \boldsymbol{\beta}_*(\tau_s)']'$, respectively. We wish to test the validity of linear restrictions on $\boldsymbol{\Phi}_*(\boldsymbol{\tau})$, $\boldsymbol{\Gamma}_*(\boldsymbol{\tau})$, and $\mathbf{B}_*(\boldsymbol{\tau})$ and express them as

$$\mathbb{H}'_0 : \mathbf{F}\boldsymbol{\Phi}_*(\boldsymbol{\tau}) = \mathbf{f} \text{ versus } \mathbb{H}'_1 : \mathbf{F}\boldsymbol{\Phi}_*(\boldsymbol{\tau}) \neq \mathbf{f}, \quad (14)$$

$$\mathbb{H}''_0 : \mathbf{S}\boldsymbol{\Gamma}_*(\boldsymbol{\tau}) = \mathbf{s} \text{ versus } \mathbb{H}''_1 : \mathbf{S}\boldsymbol{\Gamma}_*(\boldsymbol{\tau}) \neq \mathbf{s}, \quad (15)$$

$$\mathbb{H}'''_0 : \mathbf{S}\mathbf{B}_*(\boldsymbol{\tau}) = \mathbf{s} \text{ versus } \mathbb{H}'''_1 : \mathbf{S}\mathbf{B}_*(\boldsymbol{\tau}) \neq \mathbf{s}, \quad (16)$$

where \mathbf{F} and \mathbf{f} are $h \times ps$ and $h \times 1$ pre-specified matrices, respectively, and \mathbf{S} and \mathbf{s} are $h \times ks$ and $h \times 1$ pre-specified matrices with h being the number of restrictions. As a benchmark case, we test whether the short-run and the long-run parameters are equal at two quantile indices. For example, if we intend to test $\boldsymbol{\beta}_*(\tau_1) = \boldsymbol{\beta}_*(\tau_2)$, we let $\mathbf{S} = (\mathbf{I}_k, -\mathbf{I}_k)$, $\mathbf{s} = \mathbf{0}_{k \times 1}$ and $\mathbf{B}_*(\boldsymbol{\tau}) := [\boldsymbol{\beta}_*(\tau_1)', \boldsymbol{\beta}_*(\tau_2)']'$.

We impose the following assumptions in addition to Assumption 1 so that the test statistics defined below behave appropriately for moderately large sample sizes.

Assumption 2. (i) $\boldsymbol{\Xi}(\boldsymbol{\tau}) := [(f_{\tau_i} f_{\tau_j})^{-1}(\min[\tau_i, \tau_j] - \tau_i \tau_j) \mathbf{L}(\tau_i, \tau_i)^{-1} \mathbf{L}(\tau_i, \tau_j) \mathbf{L}(\tau_j, \tau_j)^{-1}]_{i,j=1,2,\dots,s}$ is a $ps \times ps$ positive-definite matrix, where $s \in \mathbb{N}$ and $\mathbf{L}(\tau_i, \tau_j) := \mathbb{E}[\widetilde{\mathbf{H}}_t(\tau_i) \widetilde{\mathbf{H}}_t(\tau_j)']$;

(ii) For each $r, \tau \in (0, 1)$, we let $\mathbf{G}_n(r, \tau) := n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} \psi_\tau[U_t(\tau)] [\widetilde{\mathbf{W}}_t', \mathbf{K}_t(\tau)']'$ and suppose that $\mathbf{G}_n(\cdot, \cdot) \Rightarrow \mathcal{G}(\cdot, \cdot)$, where $\mathcal{G}(\cdot, \cdot) := [\mathcal{B}_\psi(\cdot, \cdot), \mathcal{B}_{\psi \cdot \mathbf{W}}(\cdot, \cdot)', \mathcal{B}_{\psi \cdot \mathbf{K}}(\cdot, \cdot)']'$ is a multivariate Gaussian process such that $\mathbb{E}[\mathcal{G}(\cdot, \cdot)] = \mathbf{0}$, and for each $r, \tilde{r}, \tau, \tilde{\tau} \in (0, 1)$, $\mathbb{E}[\mathcal{G}(r, \tau) \mathcal{G}(\tilde{r}, \tilde{\tau})'] = (\min[\tau, \tilde{\tau}] - \tau \tilde{\tau}) \boldsymbol{\Pi}(r, \tilde{r}, \tau, \tilde{\tau})$ with $\boldsymbol{\Pi}(r, \tilde{r}, \tau, \tilde{\tau}) := \lim_{n \rightarrow \infty} \mathbb{E}[\mathbf{C}_n(r; \tau) \mathbf{C}_n(\tilde{r}; \tilde{\tau})']$ and $\mathbf{C}_n(r; \tau) := n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} [\widetilde{\mathbf{W}}_t', \mathbf{K}_t(\tau)']'$; and

(iii) $\boldsymbol{\Sigma}(\boldsymbol{\tau}) := \mathbf{T}(\boldsymbol{\tau}) \circ \mathbf{P}(\boldsymbol{\tau})$ is positive definite, where $\mathbf{T}(\boldsymbol{\tau}) := [\min[\tau_i, \tau_j] - \tau_i \tau_j]_{i,j=1,2,\dots,s}$ and $\mathbf{P}(\boldsymbol{\tau}) := [(f_{\tau_i} \{1 - \sum_{\ell=1}^p \phi_{\ell*}(\tau_i)\})^{-1} (f_{\tau_j} \{1 - \sum_{\ell=1}^p \phi_{\ell*}(\tau_j)\})^{-1}]_{i,j=1,2,\dots,s}$. \square

Some remarks are warranted. First, $\mathbf{L}(\cdot, \cdot)$ is introduced for notational simplicity. Second, if $\mathbf{K}_t(\cdot)$ is invariant to the level of τ , we write it as \mathbf{K}_t , and $\boldsymbol{\Xi}(\boldsymbol{\tau})$ can be simplified to $\boldsymbol{\Xi}(\boldsymbol{\tau}) = [(f_{\tau_i} f_{\tau_j})^{-1}(\min[\tau_i, \tau_j] - \tau_i \tau_j)]_{i,j=1,2,\dots,s} \otimes \{\mathbb{E}[\mathbf{K}_t \mathbf{K}_t'] - \mathbb{E}[\mathbf{K}_t \widetilde{\mathbf{W}}_t'] \mathbb{E}[\widetilde{\mathbf{W}}_t \widetilde{\mathbf{W}}_t']^{-1} \mathbb{E}[\widetilde{\mathbf{W}}_t \mathbf{K}_t']\}^{-1}$. Assumption 1(vi) also implies that $\boldsymbol{\Xi}(\boldsymbol{\tau})$ is positive definite, so that Assumption 2(i) becomes redundant for this case. Third, although it is possible to derive Assumption 2(ii) from Assumption 1 by imposing other primitive conditions, we introduce it for brevity. Fourth, for $\tau \in (0, 1)$, $\mathcal{G}(\cdot, \tau) \stackrel{d}{=} [\mathcal{B}_\psi(\cdot, \tau), \mathcal{B}_{\psi \cdot \mathbf{W}}(\cdot, \tau)', \mathcal{B}_{\psi \cdot \mathbf{K}}(\cdot, \tau)']'$ has a positive-definite covariance matrix by Assumption 1(vi). It is clear that the multivariate Gaussian process $[\mathcal{G}(\cdot, \tau_1)', \mathcal{G}(\cdot, \tau_2)', \dots, \mathcal{G}(\cdot, \tau_s)']'$ is well defined under Assumption 2. Finally, Assumption 2(iii) is imposed to derive a non-degenerate asymptotic distribution of the multi-quantile long-run estimator.

By Theorem 1, $\tilde{\Phi}_n(\boldsymbol{\tau})$ and $\tilde{\Gamma}_n(\boldsymbol{\tau})$ consistently estimate $\Phi_*(\boldsymbol{\tau})$ and $\Gamma_*(\boldsymbol{\tau})$, respectively. Hence, we first establish the asymptotic joint distribution of these short-run estimators by extending Theorem 1 into the multi-quantile framework.

Theorem 3 (Multi-Quantile Short-Run Estimators). *Under Assumptions 1 and 2,*

- (i) $\sqrt{n}(\tilde{\Phi}_n(\boldsymbol{\tau}) - \Phi_*(\boldsymbol{\tau})) \Rightarrow N[\mathbf{0}_{ps \times 1}, \Xi(\boldsymbol{\tau})]$; and
- (ii) $\sqrt{n}(\tilde{\Gamma}_n(\boldsymbol{\tau}) - \Gamma_*(\boldsymbol{\tau})) \Rightarrow N[\mathbf{0}_{ks \times 1}, \Lambda(\boldsymbol{\tau}) \Xi(\boldsymbol{\tau}) \Lambda(\boldsymbol{\tau})']$, where $\Lambda(\boldsymbol{\tau})$ is a $ks \times ps$ block diagonal matrix with s diagonal blocks $\beta_*(\tau_i) \boldsymbol{\nu}_p'$ for $i = 1, 2, \dots, s$. \square

We establish Theorem 3(i) by noting that

$$\sqrt{n}(\tilde{\Phi}_n(\tau_j) - \phi_*(\tau_j)) = f_{\tau_j}^{-1} \mathbf{L}(\tau_j, \tau_j)^{-1} \left(n^{-1/2} \sum_{t=1}^n \tilde{\mathbf{H}}_t(\tau_j) \psi_{\tau_j}[U_t(\tau_j)] \right) + o_{\mathbb{P}}(1). \quad (17)$$

This equality follows from (7). As $\Xi(\boldsymbol{\tau})$ is positive definite by Assumption 2(i), it is straightforward to apply the multivariate CLT and conclude Theorem 3(i). Next, we note that $\sqrt{n}(\tilde{\Gamma}_n(\boldsymbol{\tau}) - \Gamma_*(\boldsymbol{\tau})) = -\sqrt{n} \Lambda(\boldsymbol{\tau}) (\tilde{\Phi}_n(\boldsymbol{\tau}) - \Phi_*(\boldsymbol{\tau})) + o_{\mathbb{P}}(1)$ that is obtained by extending (9) into the multi-quantile version. The limit distribution of $\sqrt{n}(\tilde{\Gamma}_n(\boldsymbol{\tau}) - \Gamma_*(\boldsymbol{\tau}))$ is obtained from Theorem 3(i). Here, the asymptotic covariance matrix of $\sqrt{n}(\tilde{\Gamma}_n(\boldsymbol{\tau}) - \Gamma_*(\boldsymbol{\tau}))$ is not necessarily positive definite. Although $\Xi(\boldsymbol{\tau})$ is positive definite, $\text{rank}[\Lambda(\boldsymbol{\tau}) \Xi(\boldsymbol{\tau}) \Lambda(\boldsymbol{\tau})']$ is at most s . Note that $\text{rank}[\Lambda(\boldsymbol{\tau}) \Xi(\boldsymbol{\tau}) \Lambda(\boldsymbol{\tau})'] \leq \min[\text{rank}(\Lambda(\boldsymbol{\tau})), \text{rank}(\Xi(\boldsymbol{\tau}))]$ with $\text{rank}(\Lambda(\boldsymbol{\tau})) = s$ and $\text{rank}(\Xi(\boldsymbol{\tau})) = ps$. Theorem 3 is a multi-quantile generalisation of Theorem 1 and establishes that the multi-quantile short-run parameter estimators asymptotically follow the multivariate normal distribution. In the special case with $s = 1$, Theorem 1 is obtained.

We provide the asymptotic distribution of the multi-quantile long-run estimator below.

Theorem 4 (Multi-Quantile Long-Run Estimator). *Under Assumptions 1 and 2,*

- (i) $n(\tilde{\mathbf{B}}_n(\boldsymbol{\tau}) - \mathbf{B}_*(\boldsymbol{\tau})) \Rightarrow [\mathbf{I}_s \otimes (\int_0^1 \tilde{\mathbf{B}}_W(r) \tilde{\mathbf{B}}_W(r)' dr)^{-1}] \mathcal{J}_{\beta}(\boldsymbol{\tau})$, where $\mathcal{J}_{\beta}(\boldsymbol{\tau})_{ks \times 1} := [f_i^{-1}(1 - \sum_{j=1}^p \phi_{j*}(\tau_i))^{-1} \int_0^1 \tilde{\mathbf{B}}_W(r)' d\mathcal{B}_{\psi}(r, \tau_i)]'_{i=1, \dots, s}$; and
- (ii) $n(\mathbf{I}_s \otimes \mathbf{M}^{1/2})(\tilde{\mathbf{B}}_n(\boldsymbol{\tau}) - \mathbf{B}_*(\boldsymbol{\tau})) \Rightarrow N[\mathbf{0}, \Sigma(\boldsymbol{\tau}) \otimes \mathbf{I}_k]$. \square

Theorem 4 holds by Assumption 2. Here, the asymptotic covariance matrix $\Sigma(\boldsymbol{\tau}) \otimes \mathbf{I}_k$ is positive definite by Assumption 2(iii). Notice that if $k < p$, $\sqrt{n}(\tilde{\Gamma}_n(\boldsymbol{\tau}) - \Gamma_*(\boldsymbol{\tau}))$ may converge to a degenerate distribution. Nevertheless, by Assumption 2(iii), the asymptotic distribution of $n(\mathbf{I}_s \otimes \mathbf{M}^{1/2})(\tilde{\mathbf{B}}_n(\boldsymbol{\tau}) - \mathbf{B}_*(\boldsymbol{\tau}))$ is nondegenerate.

We now extend the single-quantile Wald statistics into the multi-quantile counterparts by using the

asymptotic mixture normal distribution in Theorems 3 and 4. We consider the following test statistics:

$$\begin{aligned}\mathcal{W}_n(\Phi) &:= n \left(\mathbf{F} \tilde{\Phi}_n(\tau) - \mathbf{f} \right)' \left(\mathbf{F} \hat{\Xi}_n(\tau) \mathbf{F}' \right)^{-1} \left(\mathbf{F} \tilde{\Phi}_n(\tau) - \mathbf{f} \right), \\ \mathcal{W}_n(\Gamma) &:= n \left(\mathbf{S} \tilde{\Gamma}_n(\tau) - \mathbf{s} \right)' \left(\mathbf{S} \hat{\Lambda}_n(\tau) \hat{\Xi}_n(\tau) \hat{\Lambda}_n(\tau)' \mathbf{S}' \right)^{-1} \left(\mathbf{S} \tilde{\Gamma}_n(\tau) - \mathbf{s} \right), \quad \text{and} \\ \mathcal{W}_n(\mathbf{B}) &:= n^2 \left(\mathbf{S} \tilde{\mathbf{B}}_n(\tau) - \mathbf{s} \right)' \left[\mathbf{S} \left(\hat{\Sigma}_n(\tau) \otimes \mathbf{M}^{-1} \right) \mathbf{S}' \right]^{-1} \left(\mathbf{S} \tilde{\mathbf{B}}_n(\tau) - \mathbf{s} \right)\end{aligned}$$

to test the joint hypotheses in (14), (15), and (16), respectively, where for each $i, j = 1, 2, \dots, s$,

$$\begin{aligned}\hat{\Xi}_n(\tau) &:= \left[\hat{f}_{\tau_i}^{-1} \hat{f}_{\tau_j}^{-1} (\min[\tau_i, \tau_j] - \tau_i \tau_j) \hat{\mathbf{L}}_n(\tau_i, \tau_i)^{-1} \hat{\mathbf{L}}_n(\tau_i, \tau_j) \hat{\mathbf{L}}_n(\tau_j, \tau_j)^{-1} \right]_{i,j=1,\dots,s}; \\ \hat{\mathbf{L}}_n(\tau_i, \tau_j) &:= \left(n^{-1} \hat{\mathbf{K}}(\tau_i)' \hat{\mathbf{K}}(\tau_j) \right) - \left(n^{-1} \hat{\mathbf{K}}(\tau_i)' \tilde{\mathbf{W}} \right) \left(n^{-1} \tilde{\mathbf{W}}' \tilde{\mathbf{W}} \right)^{-1} \left(n^{-1} \tilde{\mathbf{W}}' \hat{\mathbf{K}}(\tau_j) \right); \\ \hat{\Sigma}_n(\tau) &:= \left[(\min[\tau_i, \tau_j] - \tau_i \tau_j) \hat{f}_{\tau_i}^{-1} \hat{f}_{\tau_j}^{-1} \left(1 - \sum_{j=1}^p \tilde{\phi}_{j,n}(\tau_i) \right)^{-1} \left(1 - \sum_{j=1}^p \tilde{\phi}_{j,n}(\tau_j) \right)^{-1} \right]_{i,j=1,\dots,s};\end{aligned}$$

$\hat{\Lambda}_n(\tau)$ is a $ks \times ps$ block diagonal matrix with $\tilde{\beta}_n(\tau_i) \iota'_p$ as its i -th diagonal block ($i = 1, 2, \dots, s$); and \hat{f}_{τ_j} is a consistent estimator for f_{τ_j} for $j = 1, 2, \dots, s$. Here, $\hat{\Sigma}_n(\tau)$ consistently estimates $\Sigma(\tau)$ by Theorem 1, and $\hat{\mathbf{K}}(\tau)$ is assumed to consistently estimate $\mathbf{K}(\tau) := [\mathbf{K}_1(\tau), \dots, \mathbf{K}_n(\tau)]'$ in the sense that for each τ_i and τ_j , $\hat{\mathbf{L}}_n(\tau_i, \tau_j)$ consistently estimates $\mathbf{L}(\tau_i, \tau_j)$.

The null and alternative asymptotic behaviors of the multi-quantile Wald test statistics $\mathcal{W}_n(\Phi)$, $\mathcal{W}_n(\Gamma)$, and $\mathcal{W}_n(\mathbf{B})$ are summarised in the following corollary.

Corollary 4. *Suppose that Assumptions 1 and 2 hold, and for $j = 1, 2, \dots, s$, $\hat{f}_{\tau_j} \xrightarrow{\mathbb{P}} f_{\tau_j}$.*

(i) *If $\text{rank}(\mathbf{F}) = h$, and for each τ_i and τ_j , $\hat{\mathbf{L}}_n(\tau_i, \tau_j) \xrightarrow{\mathbb{P}} \mathbf{L}(\tau_i, \tau_j)$, then*

(i.a) $\mathcal{W}_n(\Phi) \Rightarrow \mathcal{X}_h^2$ under \mathbb{H}'_0 ; and

(i.b) $P(\mathcal{W}_n(\Phi) \geq d'_n) \rightarrow 1$ for any sequence $\{d'_n\}$ s.t. $d'_n = o(n)$ under \mathbb{H}'_1 ;

(ii) *If $\text{rank}(\mathbf{S} \hat{\Lambda}_n(\tau) \hat{\Xi}_n(\tau) \hat{\Lambda}_n(\tau)' \mathbf{S}') = h \leq s$, and for each τ_i and τ_j , $\hat{\mathbf{L}}_n(\tau_i, \tau_j) \xrightarrow{\mathbb{P}} \mathbf{L}(\tau_i, \tau_j)$, then*

(ii.a) $\mathcal{W}_n(\Gamma) \Rightarrow \mathcal{X}_h^2$ under \mathbb{H}''_0 ; and

(ii.b) $P(\mathcal{W}_n(\Gamma) \geq d''_n) \rightarrow 1$ for any sequence $\{d''_n\}$ s.t. $d''_n = o(n)$ under \mathbb{H}''_1 ; and

(iii) *If $\text{rank}(\mathbf{S}) = h$, then*

(iii.a) $\mathcal{W}_n(\mathbf{B}) \Rightarrow \mathcal{X}_h^2$ under \mathbb{H}'''_0 ; and

(iii.b) $P(\mathcal{W}_n(\mathbf{B}) \geq d'''_n) \rightarrow 1$ for any sequence $\{d'''_n\}$ s.t. $d'''_n = o(n^2)$ under \mathbb{H}'''_1 . \square

As Corollary 4 follows from Theorems 3 and 4, its proof is omitted. Notice that Corollary 4 imposes an additional condition, $\text{rank}[\mathbf{S} \hat{\Lambda}_n(\tau) \hat{\Xi}_n(\tau) \hat{\Lambda}_n(\tau)' \mathbf{S}'] \leq s$, as $\text{rank}[\hat{\Lambda}_n(\tau) \hat{\Xi}_n(\tau) \hat{\Lambda}_n(\tau)']$ is at most s .

4 Monte Carlo Simulations

This section conducts Monte Carlo experiments to investigate the finite sample performances of our proposed estimators and test statistics. We separate our simulations into two parts. First, we verify the theoretical claims developed in Section 3 by assuming that the QARDL orders are known. Next, we consider the case where the unknown QARDL orders are estimated by BIC. Furthermore, we compare the performance of the Wald test relative to the augmented-model based Wald test considered by Xiao (2009) using the case of unknown lag order.

4.1 Known QARDL Orders

In this subsection, we focus on the joint testing procedure under the multiple quantiles framework. To this end, we examine the following QARDL(1,1) process:

$$Y_t = \alpha_*(\tau) + \phi_*(\tau)Y_{t-1} + \theta_{0*}(\tau)X_t + \theta_{1*}(\tau)X_{t-1} + U_t(\tau) \quad \text{and} \quad X_t = X_{t-1} + W_t, \quad (18)$$

where $U_t(\tau) := U_t - F^{-1}(\tau)$, $W_t := \rho_*R_{t-1} + (1 - \rho_*^2)R_t$ with $R_t \sim \text{IID } N(0, 1)$, and U_t is generated independently of R_t . We generate U_t to follow either normal or t distribution: $U_t \sim \text{IID } N(0, 1)$ or $U_t \sim \text{IID } t_5$. Notice that t_5 distribution can deal with fat-tails. We consider three different quantile indices, $\tau_1 = 0.25$, $\tau_2 = 0.50$, and $\tau_3 = 0.75$. By construction, the true values of $\phi_*(\tau)$, $\theta_{0*}(\tau)$, and $\theta_{1*}(\tau)$ are set to be the same for each τ although $\alpha_*(\tau)$ can vary with τ .

We examine the finite sample performance of the Wald statistics using the QARDL(1,1) model. We consider the three Wald test statistics $\mathcal{W}_n(\mathbf{B})$, $\mathcal{W}_n(\Phi)$, and $\mathcal{W}_n(\Gamma)$. First, we test the following four hypotheses on the long-run parameters $\mathbf{B}_*(\tau) = [\beta_*(0.25), \beta_*(0.5), \beta_*(0.75)]'$: for $j = 1, 2, 3$, and 4,

$$\mathbb{H}_0^{(j)}(\mathbf{B}) : \mathbf{S}_j \mathbf{B}_*(\tau) = \mathbf{s}_j \quad \text{versus} \quad \mathbb{H}_1^{(j)}(\mathbf{B}) : \mathbf{S}_j \mathbf{B}_*(\tau) \neq \mathbf{s}_j, \quad (19)$$

where $\mathbf{S}_1 = [1, -1, 0]$, $\mathbf{S}_2 = [0, 1, -1]$, $\mathbf{S}_3 = [1, 0, -1]$ and $\mathbf{S}_4 = [\mathbf{S}'_1, \mathbf{S}'_2]'$ with $\mathbf{s}_1 = \mathbf{s}_2 = \mathbf{s}_3 = 0$, and $\mathbf{s}_4 = [0, 0]'$. We denote these Wald statistics respectively as $\mathcal{W}_n^{(1)}(\mathbf{B})$, $\mathcal{W}_n^{(2)}(\mathbf{B})$, $\mathcal{W}_n^{(3)}(\mathbf{B})$, and $\mathcal{W}_n^{(4)}(\mathbf{B})$. Second, we test the four restrictions on the short-run parameters, $\Phi_*(\tau) = [\phi_*(0.25), \phi_*(0.5), \phi_*(0.75)]'$ and $\Gamma_*(\tau) = [\gamma_*(0.25), \gamma_*(0.5), \gamma_*(0.75)]'$: for $j = 1, 2, 3$, and 4,

$$\mathbb{H}_0^{(j)}(\Phi) : \mathbf{S}_j \Phi_*(\tau) = \mathbf{s}_j \quad \text{versus} \quad \mathbb{H}_1^{(j)}(\Phi) : \mathbf{S}_j \Phi_*(\tau) \neq \mathbf{s}_j, \quad \text{and} \quad (20)$$

$$\mathbb{H}_0^{(j)}(\Gamma) : \mathbf{S}_j \Gamma_*(\tau) = \mathbf{s}_j \quad \text{versus} \quad \mathbb{H}_1^{(j)}(\Gamma) : \mathbf{S}_j \Gamma_*(\tau) \neq \mathbf{s}_j. \quad (21)$$

For each $j = 1, 2, 3$, and 4, these Wald statistics are denoted respectively as $\mathcal{W}_n^{(j)}(\Phi)$ and $\mathcal{W}_n^{(j)}(\Gamma)$.

To estimate the relevant statistics for testing the hypotheses in (19), (20), and (21), the density function needs to be consistently estimated. For this purpose, we employ the following kernel density estimator:

$$\hat{f}_{\tau_\ell} := \frac{1}{nh_B(\tau_\ell)} \sum_{t=1}^n \phi \left(-\frac{\hat{U}_t(\tau_\ell)}{h_B(\tau_\ell)} \right), \quad \ell = 1, 2, 3,$$

where h_B is the bandwidth proposed by Bofinger (1975): $h_B(\tau_\ell) := n^{-1/5} [4.5\phi(\Phi^{-1}(\tau_\ell))^4 / (2(\Phi^{-1}(\tau_\ell))^2 + 1)^2]^{1/5}$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal PDF and CDF, respectively. It is well established that \hat{f}_{τ_ℓ} is a consistent estimator of f_{τ_ℓ} , *e.g.*, Koenker and Xiao (2002).⁷ Next, we consistently estimate $\mathbf{\Pi}(\tau)$ for evaluating $\mathcal{W}_n^{(j)}(\Phi)$ and $\mathcal{W}_n^{(j)}(\Gamma)$. For this, (5) is first estimated by the quantile regression, and we compute $\{\hat{K}_{t,i}(\tau)\}$ as the quantile regression error. Using the sequence of $\{W_t\}$, $\mathbf{\Pi}(\tau)$ is estimated by (13).

We summarise the null simulation setup as follows: first, for each τ , we set the parameter values at $\alpha_*(\tau) = 1$, $\phi_*(\tau) = 0.25$, $\theta_{0*}(\tau) = 2$, $\theta_{1*}(\tau) = 3$, and $\rho_* = 0.5$, and we examine the empirical rejection rates of the Wald tests under the null hypotheses. By construction, $\gamma_*(\tau) = 5$ and $\beta_*(\tau) = 20/3$. We set the number of replications at 5,000 and report the empirical rejection rates of $\mathcal{W}_n^{(j)}(\mathbf{B})$, $\mathcal{W}_n^{(j)}(\Phi)$, and $\mathcal{W}_n^{(j)}(\Gamma)$ for six different sample sizes, $n = 50, 200, 400, 600, 800, 1,000$, and 2,000.

We consider the finite sample testing performance for the long-run parameter. Table 1 presents the empirical rejection frequencies of $\mathcal{W}_n^{(j)}(\mathbf{B})$ under the null hypothesis $\mathbb{H}_0^{(j)}(\mathbf{B})$ at a significance level of 5%. In all four cases, the Wald statistics tend to moderately over-reject the null when the sample size is relatively small for both normal and t errors. Especially for $n = 50$, the level distortions are non-negligible, but this is not a surprising finding given the common empirical findings in the stationary case that requires a sufficiently large sample for an accurate estimation of the quantile regression. As the sample size increases, as expected, empirical levels of all four Wald statistics tend to converge to the nominal level.

<<<<<<< Insert Table 1 around here. >>>>>>>>

These findings generally affirm the theoretical predictions in Corollary 4(iii.a) that $\mathcal{W}_n^{(j)}(\mathbf{B})$ and $\mathcal{W}_n^{(4)}(\mathbf{B})$ converge to \mathcal{X}_1^2 and \mathcal{X}_2^2 asymptotically under $\mathbb{H}_0^{(j)}(\mathbf{B})$ and $\mathbb{H}_0^{(4)}(\mathbf{B})$, respectively ($j = 1, 2, 3$).

Second, we turn to testing the performance of the short-run parameters under the null. Tables 2 and 3 present the empirical rejection rates of $\mathcal{W}_n^{(j)}(\Phi)$ and $\mathcal{W}_n^{(j)}(\Gamma)$ under $\mathbb{H}_0^{(j)}(\Phi)$ and $\mathbb{H}_0^{(j)}(\Gamma)$ at a significance level of 5%. These statistics tend to moderately over-reject the null when the sample size is relatively small.

⁷Alternatively, the bandwidth advocated by Hall and Sheather (1988) can be used to estimate the density. However, this method is affected by a particular choice of a significance level that is used as an ingredient of the estimator.

Nevertheless, as expected, empirical levels of all four Wald statistics tend to converge to the nominal levels as the sample size increases for both normal and t errors.

<<<<<<< Insert Tables 2 and 3 around here. >>>>>>>>>

These aspects are as claimed in Corollary 4(i.a, ii.a): (i) $\mathcal{W}_n^{(j)}(\Phi)$ and $\mathcal{W}_n^{(4)}(\Phi)$ converge to \mathcal{X}_1^2 and \mathcal{X}_2^2 asymptotically under $\mathbb{H}_0^{(j)}(\Phi)$ and $\mathbb{H}_0^{(4)}(\Phi)$, respectively; (ii) $\mathcal{W}_n^{(j)}(\Gamma)$ and $\mathcal{W}_n^{(4)}(\Gamma)$ converge to \mathcal{X}_1^2 and \mathcal{X}_2^2 asymptotically under $\mathbb{H}_0^{(j)}(\Gamma)$ and $\mathbb{H}_0^{(4)}(\Gamma)$, ($j = 1, 2, 3$).

Next, we report the power performance of the Wald statistics under the alternative. For this, we let $s_j = 0.1$ ($j = 1, 2, 3$) and $s_4 = [0.1, 0.1]'$ and consider the sample sizes $\{50, 100, 200, 300, 400, 500\}$ for $\mathcal{W}_n^{(j)}(\mathbf{B})$ and $\mathcal{W}_n^{(j)}(\Phi)$ and $\{200, 400, 600, 800, 1200, 1600, 2000\}$ for $\mathcal{W}_n(\Gamma)$. We summarise the simulation results as follows: first, Table 4 reports the empirical rejection rates of the $\mathcal{W}_n^{(j)}(\mathbf{B})$ tests for the long-run parameter under the four alternative hypotheses $\mathbb{H}_1^{(j)}(\mathbf{B})$ ($j = 1, 2, 3$) and $\mathbb{H}_1^{(4)}(\mathbf{B})$, from which we find that the powers increase quickly with the sample size. This is a satisfactory finding given that the alternative hypotheses are not much different from the null and that the different error distributions produce similar results. Overall, this evidence provides strong support for the theoretical predictions in Corollary 4(iii.b) that $\mathcal{W}_n^{(j)}(\mathbf{B})$ is consistent.

<<<<<<< Insert Table 4 around here. >>>>>>>>>

Second, we present the empirical rejection rates of $\mathcal{W}_n^{(j)}(\Phi)$ and $\mathcal{W}_n^{(j)}(\Gamma)$ for the short-run parameters under the alternative hypotheses, respectively, in Tables 5 and 6. The power of $\mathcal{W}_n^{(j)}(\Phi)$ approaches unity very quickly, faster than that of $\mathcal{W}_n^{(j)}(\mathbf{B})$. Surprisingly, however, we observe that the power of $\mathcal{W}_n^{(j)}(\Gamma)$ approaches unity at a much slower rate than those of $\mathcal{W}_n^{(j)}(\mathbf{B})$ or $\mathcal{W}_n^{(j)}(\Phi)$.

<<<<<<< Insert Tables 5 and 6 around here. >>>>>>>>>

Two remarks are in order. First, the finite sample performance of the Wald tests depends on the distance between quantiles. Recall that the distance between quantile indices for $\mathbb{H}_0^{(1)}(\cdot)$ and $\mathbb{H}_0^{(2)}(\cdot)$ is 0.25 (*i.e.*, $|\tau_1 - \tau_2| = |\tau_2 - \tau_3| = 0.25$), whereas the distance is 0.5 under $\mathbb{H}_0^{(3)}(\cdot)$ ($|\tau_1 - \tau_3| = 0.5$). The level distortions of $\mathcal{W}_n^{(j)}(\mathbf{B})$, $\mathcal{W}_n^{(j)}(\Phi)$, and $\mathcal{W}_n^{(j)}(\Gamma)$ are smaller for $\mathbb{H}_0^{(1)}(\cdot)$ and $\mathbb{H}_0^{(2)}(\cdot)$ than for $\mathbb{H}_0^{(3)}(\cdot)$. On the other hand, the empirical powers of $\mathcal{W}_n^{(j)}(\mathbf{B})$, $\mathcal{W}_n^{(j)}(\Phi)$, and $\mathcal{W}_n^{(j)}(\Gamma)$ are higher against $\mathbb{H}_0^{(1)}(\cdot)$ and $\mathbb{H}_0^{(2)}(\cdot)$ than against $\mathbb{H}_0^{(3)}(\cdot)$. Second, it is easily seen from Tables 4, 5, and 6 that the empirical powers of $\mathcal{W}_n^{(4)}(\mathbf{B})$, $\mathcal{W}_n^{(4)}(\Phi)$, and $\mathcal{W}_n^{(4)}(\Gamma)$ are higher than those of $\mathcal{W}_n^{(j)}(\mathbf{B})$, $\mathcal{W}_n^{(j)}(\Phi)$, and $\mathcal{W}_n^{(j)}(\Gamma)$ ($j = 1, 2, 3$). Nevertheless, this result should be somewhat discounted because the level distortions of $\mathcal{W}_n^{(4)}(\mathbf{B})$, $\mathcal{W}_n^{(4)}(\Phi)$,

and $\mathcal{W}_n^{(4)}(\Gamma)$ are also higher than those of $\mathcal{W}_n^{(j)}(\mathbf{B})$, $\mathcal{W}_n^{(j)}(\Phi)$, and $\mathcal{W}_n^{(j)}(\Gamma)$ ($j = 1, 2, 3$). This implies that testing hypotheses with a smaller quantile interval tends to deliver more precise inference results, especially when the sample size is relatively small. Alternatively, sufficiently large samples are required for reliable inference when the quantile interval is relatively wide and/or multiple restrictions are tested.

4.2 Unknown QARDL Orders

4.2.1 Testing Under the Multiple Quantile Framework

We extend the simulation set-up and consider the case with unknown QARDL orders. To this end, we construct the DGP as follows:

$$Y_t = \alpha_*(\tau) + \phi_*(\tau)Y_{t-1} + \theta_{0*}(\tau)X_t + \theta_{1*}(\tau)X_{t-1} + U_t(\tau) \quad \text{and} \quad X_t = X_{t-1} + W_t, \quad (22)$$

where $U_t(\tau) := U_t - F^{-1}(\tau)$, $U_t := \sigma_* R_{t-1}^{(1)} + (1 - \sigma_*^2) R_t^{(1)}$, and $W_t := \rho_* R_{t-1}^{(2)} + (1 - \rho_*^2) R_t^{(2)}$ with $(R_t^{(1)}, R_t^{(2)})' \sim \text{IID } N(\mathbf{0}, \mathbf{I}_2)$. Notice that the only difference of this DGP from that of (18) is that we now allow $\{U_t\}$ to be an MA(1) process with $\sigma_* \in (0, 1)$. If $\sigma_* = 0$, then the DGP condition is identical to that in (18) with normal errors. We are interested in examining how the Wald test statistics perform in the presence of the MA error terms.

In this case, we estimate the unknown QARDL orders using the BIC. Specifically, we consider 49 models by letting $(p, q) = \{(1, 1), (1, 2), \dots, (7, 6), (7, 7)\}$ and select the best fitting model with the smallest BIC. Note that the conditional mean equation is the weighted average of the quantile equations and BIC can consistently estimate the ARDL orders as confirmed by Pesaran and Shin (1998). On the other hand, applying the BIC may estimate consistently higher orders for some quantiles.

Our simulations are conducted across three quantile levels, $\tau_1 = 0.25$, $\tau_2 = 0.50$ and $\tau_3 = 0.75$ as follows: for the long-run parameters, $\mathbf{B}_*(\tau) = [\beta_*(0.25), \beta_*(0.5), \beta_*(0.75)]'$ and the short-run parameters, $\Gamma_*(\tau) := [\gamma_*(0.25), \gamma_*(0.5), \gamma_*(0.75)]'$, we use $\widetilde{\mathcal{W}}_n^{(j)}(\mathbf{B})$ and $\widetilde{\mathcal{W}}_n^{(j)}(\Gamma)$ to test the hypotheses: $\mathbb{H}_0^{(j)}(\mathbf{B}) : \mathbf{S}_j \mathbf{B}_*(\tau) = \mathbf{s}_j$ versus $\mathbb{H}_1^{(j)}(\mathbf{B}) : \mathbf{S}_j \mathbf{B}_*(\tau) \neq \mathbf{s}_j$ and $\mathbb{H}_0^{(j)}(\Gamma) : \mathbf{S}_j \Gamma_*(\tau) = \mathbf{s}_j$ versus $\mathbb{H}_1^{(j)}(\Gamma) : \mathbf{S}_j \Gamma_*(\tau) \neq \mathbf{s}_j$ for $j = 1, 2, 3, 4$. Next, for $\Phi_*(\tau) := [\phi_{1*}(0.25), \dots, \phi_{p*}(0.25), \phi_{1*}(0.50), \dots, \phi_{p*}(0.50), \phi_{1*}(0.75), \dots, \phi_{p*}(0.75)]'$, we use $\widetilde{\mathcal{W}}_n^{(j)}(\Phi)$ to test the hypotheses: $\widetilde{\mathbb{H}}_0^{(j)}(\Phi) : \widetilde{\mathbf{S}}_j \Phi_*(\tau) = \mathbf{s}_j$ versus $\widetilde{\mathbb{H}}_1^{(j)}(\Phi) : \widetilde{\mathbf{S}}_j \Phi_*(\tau) \neq \mathbf{s}_j$, for $j = 1, 2, 3, 4$, where $\widetilde{\mathbf{S}}_1 = [1, \mathbf{0}'_{p-1}, -1, \mathbf{0}'_{p-1}, 0, \mathbf{0}'_{p-1}]$, $\widetilde{\mathbf{S}}_2 = [0, \mathbf{0}'_{p-1}, 1, \mathbf{0}'_{p-1}, -1, \mathbf{0}'_{p-1}]$, $\widetilde{\mathbf{S}}_3 = [1, \mathbf{0}'_{p-1}, 0, \mathbf{0}'_{p-1}, -1, \mathbf{0}'_{p-1}]$, and $\widetilde{\mathbf{S}}_4 = [\widetilde{\mathbf{S}}_1', \widetilde{\mathbf{S}}_2']'$.

In principle, if the different lag orders are selected based on BIC, the null hypotheses can be different for the short-run parameters (of course, the same for the long-run parameter). Nevertheless, the essence

of inference is intact when testing the coefficients on the first lagged dependent variable. For example, $\widetilde{\mathcal{W}}_n^{(1)}(\Phi)$ tests $\phi_{1*}(0.25) = \phi_{1*}(0.50)$, and $\widetilde{\mathcal{W}}_n^{(4)}(\Phi)$ tests $\phi_{1*}(0.25) = \phi_{1*}(0.50) = \phi_{1*}(0.75)$. Under the DGP condition in (22), we expect the empirical rejection rates of the Wald test statistics to be close to the nominal level. For $j = 1, 2$, and 3 , $\widetilde{\mathcal{W}}_n^{(j)}(\mathbf{B})$, $\widetilde{\mathcal{W}}_n^{(j)}(\Phi)$, and $\widetilde{\mathcal{W}}_n^{(j)}(\Gamma)$ follow \mathcal{X}_1^2 ; and $\widetilde{\mathcal{W}}_n^{(4)}(\mathbf{B})$, $\widetilde{\mathcal{W}}_n^{(4)}(\Phi)$, and $\widetilde{\mathcal{W}}_n^{(4)}(\Gamma)$ follow \mathcal{X}_2^2 under the null. We consider sample sizes of 50, 200, 400, 600, 800, 1,000, and 2,000, and we let $\sigma_* = 0.00, 0.10, 0.20, 0.30$, and 0.40 . Notice that, as σ_* increases, $\{U_t\}$ becomes more serially correlated; and thus, the larger level distortion is expected for $\sigma_* = 0.4$. We compute the Wald test statistics using the same methodology as in the previous subsection and present simulation results in Tables 7, 8, and 9. For brevity, we examine the empirical rejection rates at a significance level of 5% only.

<<<<<<< Insert Tables 7, 8, and 9 around here. >>>>>>>>

The overall simulation results are as follows. First, all the empirical levels of the Wald tests tend to the nominal level as the sample size increases, implying that the Wald tests are asymptotically well approximated by the claimed null distributions even in the presence of MA(1) errors. Second, as expected, there are somewhat non-negligible level distortions in the small sample (say, $n = 50$) as σ_* increases. Hence, the researcher needs to be cautious when using the proposed test with small sample sized data. Nevertheless, for all the values of σ_* , they are almost negligible as the sample size becomes moderately large.

4.2.2 Comparison with the Wald Test in Xiao (2009)

In this subsection we aim to compare the finite sample performance of our proposed Wald tests relative to that of Xiao's (2009) augmented model-based Wald test. The augmented quantile model in Xiao (2009) is specified as: $Q_{Y_t}(\tau|\mathbf{Z}_t) = \alpha_*(\tau) + \mathbf{X}'_t\beta_*(\tau) + \sum_{j=-K}^K \Delta\mathbf{X}'_{t-j}\mathbf{\Pi}_*(\tau)$. Accordingly, the augmented Wald statistic is obtained by $\mathfrak{W}_n^{(K)}(\beta) := \frac{\hat{f}_\tau^2}{\tau(1-\tau)}(\mathbf{R}\widehat{\beta}_n(\tau) - \mathbf{r})'[\mathbf{R}\mathbf{M}_\mathbf{X}^{-1}\mathbf{R}]^{-1}(\mathbf{R}\widehat{\beta}_n(\tau) - \mathbf{r})$, where $\mathbf{M}_\mathbf{X} := \sum_{t=1}^n(\mathbf{X}_t - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})'$ and $\bar{\mathbf{X}}$ is the sample average of $\{\mathbf{X}_t\}$. It tests the restrictions on the long-run parameters: $\mathbb{H}_0^* : \mathbf{R}\beta_*(\tau) = \mathbf{r}$ vs $\mathbb{H}_1^* : \mathbf{R}\beta_*(\tau) \neq \mathbf{r}$. For sufficiently large K , Xiao (2009) shows that $\mathfrak{W}_n^{(K)}(\beta)$ follows \mathcal{X}^2 distribution asymptotically under his regularity conditions.

To make the fair comparison, we set up the simulation design as follows: First, we employ the DGP with unknown lag orders as considered in Section 4.2.1 and test the hypotheses on the long-run parameters: $\mathbb{H}_0^* : \beta_*(\tau) = \beta_0$ vs. $\mathbb{H}_1^* : \beta_*(\tau) \neq \beta_0$, where τ is set at 0.5 for simplicity and we set $\beta_0 = (\theta_{0*} + \theta_{1*})/(1 - \phi_*)$ and $(\theta_{0*} + \theta_{1*})/(1 - \phi_*) + 0.1$ under the null and under the alternative hypothesis, respectively. Other simulation set-ups and the testing procedures for the Wald tests are identical to those in Section 4.2.1. Second, we only consider the case with $\phi_* = 0$. If $\phi_* \neq 0$, then the DGP does not satisfy Assumption A'

in Xiao (2009) because the value of K should be infinite.⁸ Third, as unknown QARDL orders are selected by BIC in Section 4.2.1, we have also selected unknown K by BIC. Specifically, we select the best fitting model with the smallest BIC for $K = 1, 2, \dots, 7$.

These simulation results are reported in Table 10. First, as described in 4.2.1, the rejection frequencies of our proposed Wald test tends to the nominal level (5%) quickly as the sample size rises, despite the serial correlation in U_t . By contrast, Xiao’s $\mathfrak{W}_n^{(K)}(\beta)$ test tends to display large size distortions. Only in the case where the errors are not serially correlated ($\sigma_* = 0$), its empirical rejection frequency becomes close to the nominal level as the sample size increases. Next, turning to the (size-adjusted) power performance, we find that both tests tend to display similar power across all the sample sizes considered. Overall, the simulation results clearly show potential advantages of our proposed testing procedure, especially in the context of most time series data being subject to complex dynamics.

<<<<<<< Insert Table 10 around here. >>>>>>>>

5 Empirical Application

In a classic study on dividend policy, based on interviews with 28 managers, Lintner (1956) observes that firms gradually adjust dividends in response to changes in earnings. Lintner also observes that firms are reluctant to make dividend changes that have to be reversed in the near future. An important implication of this finding is that managers make dividend adjustments in response to unanticipated and non-transitory changes in their firms’ earnings in order to attain a long-run target payout ratio. Empirical research generally supports Lintner’s partial adjustment framework at both the firm and the aggregate level, *e.g.*, Fama and Babiak (1968), Marsh and Merton (1987), Garrett and Priestley (2000), and Andres *et al.* (2009). In the literature, however, the estimates of adjustment speed are widely different. Using the US real estate investment trusts data from 1992-2003, Hanyunga and Stephens (2009) report widely different estimates of the speed of adjustment: 0.028 and 0.04 at a quarterly frequency and 0.38 and 0.37 at an annual frequency by the panel OLS and Tobit methods, respectively. These large differences are mainly due to the well-known small- T bias in dynamic AR regression (*e.g.*, Hurwicz, 1950).

Recently, Brav *et al.* (2005) surveyed 384 financial executives to determine the factors that drive dividend and share repurchase decisions. The most important findings are as follows: executives try to avoid reducing dividends per share (93.8% agreed) and aims to maintain a smooth dividend stream (89.6%); they

⁸We have also conducted an additional simulation experiment using the same DGP but with $\phi_* = 0.25$. To save space we do not report these simulation results. Here we find that our proposed Wald test displays more or less similar results to those with $\phi_* = 0.00$ whilst the Xiao’s Wald test now tends to suffer from severe size distortions.

are reluctant to make changes that will be reversed (77.9%) because there are negative consequences to cutting dividends (88.1%). These responses are mostly consistent with Lintner's observations, although the link between dividends and earnings has weakened in the subsequent 50 years. With the new survey evidence in mind, Leary and Michaely (2011) address an important issue of why firms smooth dividends by studying the cross-sectional differences in the behavior of dividend smoothing. Their main findings suggest that dividend smoothing is most common in the U.S. in large and mature firms that are not financially constrained, possess low levels of asymmetric information, and are readily susceptible to agency conflicts. Furthermore, they document that dividend smoothing has steadily increased over the past century, even before firms began to repurchase shares in the mid-1980s. Chen *et al.* (2012) also show that aggregate dividends are dramatically more smoothed in the post-world war II period (1946-2006) than in the prewar period (1871-1945). In particular, they report that, during the post-war period, dividends adjust to the earnings target at a speed about one-fourth of the prewar period. This implies that dividend growth is unpredictable for the firms that have most smoothed dividends but predictable for the firms that have least smoothed dividends. By a simulation analysis, they draw two important conclusions. First, even if dividends are supposed to be strongly predictable without smoothing, dividend smoothing can negate this predictability. Second, dividend smoothing leads to a persistent dividend yield. Motivated by these, Rangvid *et al.* (2012) investigate the relationship between dividend predictability and dividend smoothing in a global framework using aggregated data of dividend yields, prices, and total returns in 50 countries at a quarterly frequency. They find that dividend predictability is driven by cross-country differences in firm characteristics, dividend smoothing, and institutions. In particular, aggregate dividend growth is highly predictable by dividend yields in medium-sized and smaller countries, although this predictability disappears for large countries.

However, all of these empirical findings are obtained from examining dividend behaviors only at the conditional mean, and the literature has not yet investigated an important possibility that the degree of dividend smoothing can be fundamentally heterogeneous across different quantiles of dividend distribution. The relationship between the dependent variable and covariates may differ depending on the location of the dependent variable in its own distribution. We contribute to the existing literature on dividend policy by incorporating location asymmetries into dividend adjustment and target payout ratio at the aggregate level.

Lintner (1956) suggests that firms partially adjust toward their target dividend, and models this adjustment by $\Delta D_t = a_* - \zeta_* (D_t^* - D_{t-1}) + \epsilon_t$, where D_t and D_t^* are the current (observed) and target levels of dividends at time t , respectively, and $|\zeta_*|$ is the speed of adjustment or smoothing coefficient (expected to lie between 0 and 1), which represents how quickly firms adjust toward the target dividend. It is widely

suggested that the target dividend has a long-run relationship with the current earning E_t as⁹ $D_t^* = \beta_* E_t$, where β_* is the target payout ratio. Combining this with Lintner's (1956) specification yields the following partial adjustment model: $\Delta D_t = a_* + \zeta_* D_{t-1} + \theta_* E_t + \epsilon_t$, where $\theta_* := -\zeta_* \beta_*$.

We find that D_t and E_t are non-stationary, as in Garrett and Priestley (2000). Hence, the target dividend relationship in $D_t^* = \beta_* E_t$ is indeed a cointegrating relationship between D_t and E_t (these are confirmed by unit root and the cointegration test results, which are unreported to save space). Importantly, ϵ_t is highly likely to be serially correlated. Therefore, it is more useful to generalise the model as follows: $\Delta D_t = \alpha_* + \zeta_* D_{t-1} + \gamma_* E_{t-1} + \sum_{j=1}^{p-1} \lambda_{j*} \Delta D_{t-j} + \sum_{j=0}^{q-1} d_{j*} \Delta E_{t-j} + \epsilon_t$, where ϵ_t is assumed to be serially uncorrelated given sufficiently large lag orders of p and q . Notice, however, that ΔE_t may be contemporaneously correlated with the error term ϵ_t , in which case we can control for such correlation by employing a projection of ϵ_t on ΔE_t : $\epsilon_t = \omega_* \Delta E_t + U_t$, where U_t is not correlated with ΔE_t by construction. Substituting this projection, we obtain the final ARDL(p, q) specification as

$$\Delta D_t = \alpha_* + \zeta_* (D_{t-1} - \beta_* E_{t-1}) + \sum_{j=1}^{p-1} \lambda_{j*} \Delta D_{t-j} + \sum_{j=0}^{q-1} \delta_{j*} \Delta E_{t-j} + U_t, \quad (23)$$

where $\beta_* := -\frac{\gamma_*}{\zeta_*}$, $\delta_{0*} := d_{0*} + \omega_*$ and $\delta_{j*} := d_{j*}$ for $j = 1, \dots, q-1$. Unless $\omega_* = 0$,¹⁰ the ARDL(p, q) model in (23) is unable to identify a contemporaneous causal relationship between ΔD_t and ΔE_t due to the projection between them. However, under the additional and acceptable assumption that earning changes can immediately cause dividend changes, but not vice versa, we can interpret the new coefficient on ΔE_t , δ_{0*} , as an impact reaction parameter. Furthermore, all the other parameters, including β_* and ζ_* in (23), can be estimated free of endogeneity since ΔE_t is included in the regression.

In particular, we are interested in the following four parameters: (i) the ECM parameter ζ_* measuring the degree of dividend smoothing, (ii) the long-run (cointegrating) target payout ratio β_* , (iii) the momentum effect of the dividend growth captured by $\lambda_* := \sum_{j=1}^{p-1} \lambda_{j*}$, which measures the cumulative impact of past dividend growth on the current dividend growth, *i.e.* $\sum_{j=1}^{p-1} \partial \Delta D_t / \partial \Delta D_{t-j}$ (there is pervasive evidence that stock returns with positive momentum in the short-run are followed by reversals in the long-run, *e.g.*, Koijen *et al.* (2009)), and (iv) the impact reaction of dividend growth to earnings growth, δ_{0*} , which measures the effect of contemporaneous change in earnings on the current dividend growth, *i.e.*, $\partial \Delta D_t / \partial \Delta E_t$. Based on

⁹The literature provides a number of ways to measure permanent earnings. Marsh and Merton (1987) suggest using the one-period lagged real stock price. Garrett and Priestley (2000) apply the Kalman filtering method and separate permanent earnings from a transitory component in the reported earnings. Alternatively, Andres *et al.* (2009) suggest using cash flow as an alternative measure of (permanent) earnings. In this study, we consider aggregated earnings in its simplest form.

¹⁰The sign of ω_* may dictate whether the impact effect, δ_{0*} in (23) is overreacting ($\omega_* > 0$) or underreacting ($\omega_* < 0$).

existing empirical and theoretical studies (*e.g.*, Kojien *et al.*, 2009; Chen *et al.*, 2012), our priors about these four parameters are given as follows: $\zeta_* < 0$, $0 < \beta_* < 1$, $\lambda_* \geq 0$, and $\delta_{0*} \geq 0$.

We employ the dataset collected by Shiller (2005) for Irrational Exuberance. We construct a quarterly dataset on real price, real dividend, and real earnings for the Standard and Poor's 500 stocks over the period 1871Q3 - 2010Q2 with 558 quarterly observations. As a baseline case, we first estimate the ARDL(p , q) model in (23) by OLS. Employing the BIC lag selection procedure, we find that the appropriate lag orders are $p = 3$ and $q = 1$, respectively. The conditional mean estimation results for the four key parameters are reported in Table 11(a). The ECM coefficient is -0.04 , which implies that the adjustment speed is only about 4%, whereas the long-run payout coefficient is 0.36. Further, the momentum and the impact reaction parameters are estimated at 0.48 and 0.01, respectively. Combined together, we conclude that the changes in dividend appear to be driven more strongly by history than the changes in earnings. The magnitude and the sign of all four key coefficients are generally in line with our priors.

Next, in order to apply the QARDL approach, we consider the quantile counterpart of the ARDL(3,1) model as follows:¹¹ $\Delta D_t = \alpha_*(\tau) + \zeta_*(\tau) D_{t-1} + \gamma_*(\tau) E_{t-1} + \sum_{j=1}^2 \lambda_{j*}(\tau) \Delta D_{t-j} + \delta_*(\tau) \Delta E_t + U_t(\tau)$ for $\tau \in (0, 1)$. To the best of our knowledge, our dynamic quantile regression application to an investigation of the long-run target payout policy and the associated dividend smoothing is the first attempt in the literature. Based on the existing literature discussed above, we aim to address the following issues that may challenge many studies relying upon the conditional mean models:

- *Issue 1:* locational asymmetry is associated with the notion that the four parameters may depend on the location of the dividend within its conditional distribution. In particular, we are keen to determine whether the long-run target payout ratio and dividend smoothing are heterogeneous across quantiles. This allows for a natural distinction among firms with low- and high-dividend payout policies.
- *Issue 2:* through employing the rolling estimation technique, we wish to investigate the time-varying patterns of dividend policy. Given the empirical evidence that the link between dividends and earnings has recently weakened (Brav *et al.*, 2005) and that aggregate dividends are dramatically more smoothed in the post-world war II period than in the prewar period (Chen *et al.* 2012), we are interested in elucidating whether the location asymmetries are monotonic over the whole periods.

¹¹When estimating the QARDL model, we acknowledge that the lag orders, p and q , should be selected according to quantile in a data-coherent manner. However, extending the analysis to allow for lag order selection or approximability of infinite order models could lead to an overly long and complicated paper and is beyond the scope of the current study. We have briefly addressed this issue in the Monte Carlo section using BIC. Making p and q dependent on the quantile index τ not only increases the computation complexity, but also reduces the comparability across different quantile estimation results. Thus, we find it reasonable to employ the common lag orders selected via the BIC applied to the conditional mean model across quantiles in order to make quantile estimation results comparable to the baseline conditional mean model, as similarly implemented in Covas *et al.* (2012).

The full sample quantile estimation results are reported in Table 11(b), showing that most of the coefficients are statistically significant across all of the quantiles that we consider. The only exceptions are the estimates of $\delta_*(\tau)$ at higher quantiles.

<<<<<<< Insert Table 11 around here. >>>>>>>>>

We also plot the estimation results in Figure 1, which displays the quantile estimates of the four parameters $\zeta_*(\tau)$, $\beta_*(\tau)$, $\lambda_*(\tau) = \sum_{j=1}^2 \lambda_{j*}(\tau)$, and $\delta_*(\tau)$ with 90% confidence intervals against quantile indices ranging from 0.05 to 0.95.¹²

<<<<<<< Insert Figure 1 around here. >>>>>>>>>

The estimation results are summarised as follows. First, the quantile estimates of the ECM parameter $|\widehat{\zeta}(\tau)|$ start with a 6% adjustment speed at the low quantiles ($\tau = 0.05$ and 0.1) and decrease monotonically as the quantile increases. The estimate reaches a minimum of 3% at $\tau = 0.4$ and stays between 3% and 4% at higher quantiles, which indicates that dividend smoothing is stronger in medium-to-higher quantiles than in lower quantiles; *i.e.*, the dividend policy is more conservative at medium-to-higher quantiles. This evidence is consistent with the hypotheses of the free cash flow problem and/or agency conflicts resulting from market frictions. For example, Easterbrook (1984) and DeAngelo and DeAngelo (2007) predict a positive relationship between the level of dividends and smoothing. Second, the quantile estimates of the long-run target payout ratio, $\widehat{\beta}(\tau)$ increase monotonically with quantiles, reaching a peak of 0.40 at $\tau = 0.3$ and remaining at similar levels at higher quantiles, except the highest quantile, $\tau = 0.95$ where the long-run payout ratio drops to 0.33. Such stable and high levels contrast with the low long-run payout ratio observed at low quantiles (*e.g.*, $\widehat{\beta}(\tau) = 0.27$ at $\tau = 0.05$).¹³ Third, the quantile estimates of the momentum parameter $\widehat{\lambda}(\tau)$ generally decrease with quantile indices (from 0.58 at $\tau = 0.05$ to 0.30 at $\tau = 0.9$), providing evidence of locational asymmetry. Thus, the higher momentum effects are observed at the lower quantiles. Finally, the impact reaction coefficient $\widehat{\delta}(\tau)$ tends to decrease with quantiles, although the magnitudes are negligibly small across all quantiles, ranging between 0.02 and -0.01 . This suggests that a change in current earnings flow that is viewed by management as essentially transitory would be unlikely to give rise to a noticeable and immediate change in dividends.

¹²We employed the wild bootstrap method proposed by Feng *et al.* (2011) to produce confidence intervals because bootstrapping can provide a better approximation of the underlying sampling distribution than the asymptotic theory.

¹³Notice that the difference between $\widehat{\beta}(\tau)$'s mostly stays within the confidence band except at the lower quantiles, indicating that the overall locational asymmetry is statistically weak. Similar remarks apply to other parameters whenever their differences are inside the confidence band.

The full sample estimation results clearly confirm that there is evidence of location asymmetries between lower and medium-to-higher quantiles for all four key parameters. Overall, these findings are consistent with the cross-sectional evidence by Leary and Michaely (2011) and the aggregate time series evidence shown in a global setting by Rangvid *et al.* (2012) that large and mature firms with stable cash-flow and return processes have a greater tendency to smooth dividends. Our findings of the quantile-dependent cointegrating relationship between dividends and earnings provide further support for the number of recent studies that report similar results; namely, the quantile-varying cointegrating relationship between stock price and dividend in Xiao (2009), between spot and future oil prices in Lee and Zeng (2011), and between the nominal interest rate and inflation in Tsong and Lee (2013). As discussed by Xiao (2009), a plausible explanation for quantile-varying cointegration is that the underlying relationship between integrated time series may vary over time due to (heterogeneous) shocks arising at each point of time. If so, the quantile cointegrating framework is naturally fitted for such a situation because quantile coefficients can be viewed as random coefficients, as explained in Koenker and Xiao (2006), in which such randomness is driven by a common shock arriving at each time period.

Considering that the sample period is quite long at more than 140 years, we find it more prudent to allow for time-varying patterns of dividend policy. Any model estimated over a long span of time that does not incorporate structures of time-varying mechanisms can result in only the average tendency of the dividend policy as examined using a range of regime-switching models. In this paper, we instead employ a robust rolling estimation technique with a window length of 320 quarters, a figure that should balance the data requirement of the QARDL model with our desire to examine the richest possible range of regimes. To this end, we re-estimate the QARDL(3,1) model by successively moving the estimation window forward by one quarter until we reach the end of the sample. In Figure 2 we present the time-series plots of the rolling quantile estimates of the four parameters together with 90% confidence intervals.

<<<<<<< Insert Figure 2 around here. >>>>>>>>>

The rolling quantile estimates of $|\zeta_*(\tau)|$ display quite strong time-varying patterns. In general, we observe that dividend smoothing has been stronger in recent periods, a finding consistent with Leary and Michaely (2011) and Chen *et al.* (2012). In the earlier periods, the location asymmetry was stronger with an order $|\widehat{\zeta}(0.25)| > |\widehat{\zeta}(0.5)| > |\widehat{\zeta}(0.75)|$, where the average adjustment speeds were about 8.50% and 4.00% at $\tau = 0.25$ and $\tau = 0.75$, respectively. This ordering is generally consistent with the full sample results. On the other hand, in the later periods, the location asymmetry was much weaker. Interestingly, we observe a reversed order with $|\widehat{\zeta}(0.75)| > |\widehat{\zeta}(0.5)| > |\widehat{\zeta}(0.25)|$ at some later periods, although their differences are

negligible and insignificant. In particular, we find that the speed of adjustment at lower quantiles begins to fall with a steep downward trend. For example, the adjustment speed at $\tau = 0.25$ decreases dramatically from about 9% in the earlier period to below 1% in the recent period (the late 2000s). In contrast, the adjustment speed seems quite stable over the whole period at the higher quantile $\tau = 0.75$.

Next, the rolling quantile estimates of $\beta_*(\tau)$ display a strong downward time-varying pattern, showing that payout ratios have become significantly lower in recent periods. This finding generally supports the study by Fama and French (2001) who report a substantial decline in the proportion of U.S. firms paying dividends. A significant decrease in the residual propensity to pay dividends was observed even after controlling for firm characteristics. Chen *et al.* (2012) also document that the aggregate payout ratio has declined in the postwar period. Notice, however, that this (monotonic) downward trend is observed only at $\tau = 0.5$ and 0.75 . The payout ratio at $\tau = 0.25$ tends to increase from the middle of the sample period (approximately after the 1960s) and starts to decrease from the 1970s. The location asymmetry was stronger in the earlier periods with an order $\hat{\beta}(0.75) > \hat{\beta}(0.5) > \hat{\beta}(0.25)$, where the average payout ratios are about 35% and 60% at $\tau = 0.25$ and $\tau = 0.75$, respectively. On the other hand, we observe weaker location asymmetry in the later periods but with a reversed order $\hat{\beta}(0.25) > \hat{\beta}(0.5) > \hat{\beta}(0.75)$. Especially since 1980s, the location asymmetry was almost negligible, and the payout ratio remains generally at a low level, except at the noted spikes.¹⁴ This finding can be explained by the evidence provided by Skinner (2008), who observes that repurchases are increasingly responsive to earnings by substituting themselves for dividends, especially over the last two decades, since stock repurchases emerged as significant in the early 1980s.¹⁵ Hence, the most recent ongoing downward trend can be attributed to the popularity of repurchase as a means of disbursing temporal increases of cash flow.

Turning to the rolling quantile estimates of $\lambda_*(\tau)$, we observe volatile time-varying patterns. The location asymmetry is stronger in the earlier periods with an order $\hat{\lambda}(0.25) > \hat{\lambda}(0.5) > \hat{\lambda}(0.75)$, in which the average momentum coefficients are 0.45 and 0.32 at $\tau = 0.25$ and $\tau = 0.75$, respectively. The location asymmetry is less significant in the later periods, although they tend to increase substantially at $\tau = 0.75$ (from 0.30 to 0.60) such that significant location asymmetry is observed very recently with $\hat{\lambda}(0.75) > \hat{\lambda}(0.5) \simeq \hat{\lambda}(0.25)$. From this evidence, we conclude that dividend policy has evolved in such a way that its own lags have become the more important predictor in the postwar period. Using the

¹⁴Three conspicuous spikes are observed at $\tau = 0.25$ during the late 2000s. These coincide with the corresponding three spikes in ECM parameter, where estimates of $\hat{\zeta}(0.25)$ are very close to zero. As the long-run coefficient is evaluated by $\hat{\beta}(0.25) = -\hat{\gamma}(0.25) / \hat{\zeta}(0.25)$, we should bear in mind that these three extreme estimates may be unreliably inflated.

¹⁵One of the main reasons behind this structural shift is Rule 10b-18 introduced in 1982 which provided a non-exclusive safe harbor for issuer repurchase. Furthermore, the Securities Exchange Commission proposed amendments of the rule in 2010 to clarify and modernise the safe harbor provision in light of developments in automated trading systems and technology.

univariate regression of dividend change on its own lags, Chen *et al.* (2012) also document a similar finding that the autoregressive coefficient is 0.061 for the prewar period (statistically insignificant), whereas it is 0.687 in the postwar period (significant).

Finally, the rolling quantile estimates of $\delta_*(\tau)$ show that there is a strong downward trend from 12–14% in the earliest sample period to almost zero in very recent period. These time-varying patterns are somewhat similar to those observed from $|\widehat{\zeta}(\tau)|$, and both contribute to the extremely strong dividend smoothing reported in more recent periods. The location asymmetry looks insignificant over the whole rolling period, with no clear pattern or order. In the earlier periods (mostly in the prewar period), managers tend to make a noticeable change in dividends immediately following current earnings changes, although such impact reactions decrease monotonically over time. These downward trends in conjunction with a zero bound in the most recent periods clearly confirm that managers now favor repurchase out of current earnings changes because they are more flexible than dividends, as documented in the survey evidence by Brav *et al.* (2005).

As our last step in data examination, we provide formal testing results for the location asymmetries for all four parameters over three selected quantiles $\tau = 0.25, 0.5$ and 0.75 ; *i.e.* we wish to test if each parameter is constant across quantiles. For example, if $\beta_*(\tau)$ is the parameter of interest, then we consider the following four null hypotheses: $H_{01}^\beta : \beta_*(0.25) = \beta_*(0.5)$, $H_{02}^\beta : \beta_*(0.5) = \beta_*(0.75)$, $H_{03}^\beta : \beta_*(0.25) = \beta_*(0.75)$, and $H_{04}^\beta : \beta_*(0.25) = \beta_*(0.5) = \beta_*(0.75)$. To this end, we employ the Wald test statistics proposed in Section 3, denoted as $\mathcal{W}_n^{(1)}(\beta)$, $\mathcal{W}_n^{(2)}(\beta)$, $\mathcal{W}_n^{(3)}(\beta)$, and $\mathcal{W}_n^{(4)}(\beta)$. Figure 3 plots the p -values of these Wald statistics using rolling estimation. In the earlier periods, the null hypotheses are strongly rejected across all three quantiles. In later periods, such location asymmetries become statistically significant only between $\beta_*(0.5)$ and $\beta_*(0.75)$ and between $\beta_*(0.25)$ and $\beta_*(0.75)$. Figure 4 displays the testing result for $\zeta_*(\tau)$. We find that the locational asymmetries are significantly observed only in the earlier periods, especially between $\zeta_*(0.25)$ and $\zeta_*(0.75)$. On the other hand, in the later periods, such location asymmetries become statistically insignificant, except in a few recent periods. These results confirm our earlier findings that strong locational asymmetries are observed in both target payout ratio and smoothing pattern between the low and the high quantiles only in the earlier periods.

<<<<<<< Insert Figures 3 and 4 around here. >>>>>>>>

Figures 5 and 6 provide the test results for $\lambda_*(\tau)$ and $\delta_*(\tau)$, respectively. The p -values for $\lambda_*(\tau)$ indicate that there is weak evidence in favor of the location asymmetry between $\lambda_*(0.25)$ and $\lambda_*(0.75)$ only in the very early periods. With regard to $\delta_*(\tau)$, we find that there exist some evidence of location asymmetry across quantiles around the 1990s.

<<<<<<< Insert Figures 5 and 6 around here. >>>>>>>>

In summary, we conclude that a thorough examination using the QARDL method sheds further light on the evolution of dividend policy in the U.S. over the past century. First, we document that dividend smoothing has become monotonically stronger over time. Similar monotonic downward trending patterns have been observed in the impact coefficient with respect to changes in current earnings, and they reach almost zero in the most recent periods. Both factors contribute to the extremely strong dividend smoothing reported in very recent periods. Second, our results clearly display that payout ratios have been monotonically decreasing over time and have recently stayed below 30%, providing support for the survey evidence by Brav *et al.* (2005) that the target payout ratio may no longer be the preeminent decision variable affecting payout decisions. Furthermore, we find that the location asymmetries across different quantiles of the conditional distribution of dividends, which are clearly visible in the earlier periods, are mostly negligible in the recent periods. These findings may indicate the establishment of financial deepening as a consequence of the long-term process to promote the stability of the whole financial system in the U.S. Rangvid *et al.* (2012) also provide international evidence that developed countries with more stable returns and dividend processes and with a higher quality of legal systems and corporate governance, such as the U.S., the UK, and Japan, tend to smooth dividends more than other countries.

6 Concluding Remarks

Recently, the literature on quantile time series regression, especially with nonstationary variables, has been rapidly increasing. In particular, Xiao (2009) advances a novel quantile cointegration estimation technique in a static regression. In this paper, we propose the dynamic QARDL modelling approach to simultaneously address the long-run relationship between integrated time series as well as the associated short-run dynamics across a range of quantiles of the conditional distribution of the dependent variable.

We have derived the asymptotic theory for the QARDL model with nonstationary regressors. The QARDL estimators of both the short-run dynamic and the long-run (cointegrating) parameters are shown to asymptotically follow a (mixture) normal distribution. Hence, the null distribution of the Wald statistics for testing the restrictions on both parameters weakly converges to a chi-squared distribution. We have also provided a general package in which the model can be estimated across multiple quantiles, and any linear restrictions on the parameters involving multiple quantiles can be tested using the standard Wald statistics. Monte Carlo simulation results provide strong support for theoretical predictions. The key strengths of the QARDL framework have been demonstrated in the empirical analysis of dividend policy in the U.S. using

quarterly data on real dividends and earnings for the S&P 500 stocks over the period 1871Q3 - 2010Q2.

Finally, we note several avenues for further researches. First, foremost among these must be the development of a formal cointegration testing mechanism that could address the issues of whether the long-run cointegration relationship exists at each of quantiles. In a similar context, Koenker and Xiao (2004) and Galvao (2009) suggest that the stationary property of individual time series may change across quantiles. Considering that the distribution of nonstationary variables is changing over time, it is a challenging issue to analyse such stochastic trends of nonstationary data within the scope of the quantile regression. In this regard, we conjecture that a quantile-dependent cointegration testing framework can be developed by extending the findings of Pesaran *et al.* (2001), although this approach significantly differs from the existing approaches adopted by Koenker and Xiao (2004, 2006) and Xiao (2009). Second, we find it quite useful to develop a quantile regression extension of the asymmetric ARDL framework advanced by Shin, Yu, and Greenwood-Nimmo (2014). This combined approach is expected to provide us the flexible econometric framework, which can help us to identify several forms of distinct asymmetry. Finally, given that conventional estimation can be significantly affected by conditional heteroskedasticity, it would be desirable to explicitly control for time-varying volatilities in the QARDL framework.

Acknowledgements

We are mostly grateful to the Co-editor, Han Hong, and two anonymous referees for their helpful comments. We are also grateful to Youngsoo Bae, Stephane Bonhomme, Yongsung Chang, Kausik Chaudhuri, Seonghoon Cho, In Choi, Kyungwook Choi, Viet Anh Dang, Ana-Maria Fuertes, Matthew Greenwood-Nimmo, Hwankoo Kang, Dowan Kim, Junhan Kim, Gary Koop, Dongjin Lee, Viet Hoang Ngyuen, Sangsoo Park, Yangsoo Park, Peter Phillips, Kevin Reilly, Xin Shen, Kyulee Shin, Peter Smith, Seungjoo Song, Peter Spencer, Till van Treeck, Mike Wickens, Ralf Wilke, seminar participants at the Bank of Korea, the Universities of Brunel, Korea, Leeds, Melbourne, Seoul, Sogang, Yonsei, York, Cass Business School, and the Institut für Makroökonomie und Konjunkturforschung (IMK, Dusseldorf), and conference delegates at the Conference in Honour of Professor P.C.B. Phillips (University of York, 12-13 July 2012) for their helpful comments. Kim is grateful for financial support from the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013S1A3A2053799).

References

- Acharya, V., L.H. Pedersen, T. Philippon and M. Richardson, 2011, Measuring systemic risk, Working paper, New York university.
- Adrian, T. and M. Brunnermeier, 2010, CoVaR, Working paper, Princeton university.
- Andres, C., A. Betzer, M. Goergen and L. Renneboog, 2009, Dividend policy of German firms: a panel data analysis of partial adjustment models, *Journal of empirical finance* 16, 175–187.
- Barnes, M.L. and A.W. Hughes, 2002, A quantile regression analysis of the cross section of stock market returns, Working Paper No. 02-2, Federal reserve bank of Boston, Boston, MA.
- Bofinger, E., 1975, Estimation of a density function using order statistics, *Australian journal of statistics* 17, 1–7.
- Brav, A., J. Graham, C. Harvey and R. Michaely, 2005, Payout policy in the 21st century, *Journal of financial economics* 77, 483–527.
- Burdekin, R.C.K. and P.L. Siklos, 2012, Enter the dragon: interactions between Chinese, US and Asia-Pacific equity markets, 1995–2010, *Pacific-basin finance journal* 20, 521–541.
- Chen, L., Z. Da and R. Priestley, 2012, Dividend smoothing and predictability, *Management science* 58, 1834–1853.
- Cho, J.S., C. Han and P.C.B. Phillips, 2010, LAD asymptotics for conditionally heteroskedastic time-series data with possibly infinite error densities, *Econometric theory* 26, 953–962.
- Covas, F., B. Rump and E. Zakrajsek, 2012, Stress-testing U.S. bank holding companies: a dynamic panel quantile regression approach, Discussion paper, Federal reserve board.
- DeAngelo, H. and L. DeAngelo, 2007, Payout policy pedagogy: what matters and why, *European financial management* 13, 11–27.
- Denis, D.J. and I. Osobov, 2008, Why do firms pay dividends? International evidence on the determinants of dividend Policy, *Journal of financial economics* 89, 62–82.
- Easterbrook, F., 1984, Two agency-cost explanations of dividends, *American economic review* 74, 650–659.

- Engle, R. and C.W.J. Granger, 1987, Co-integration and error Correction: representation, estimation, and testing, *Econometrica* 55, 251–276.
- Engle, R. and S. Manganelli, 2004, CAViaR: conditional autoregressive value at risk by regression quantiles, *Journal of business & economic statistics* 22, 367–381.
- Fama, E.F. and H. Blasiak, 1968, Dividend policy: an empirical analysis, *Journal of the American statistical association* 63, 1132–1161.
- Fama, E.F. and K. French, 2001, Disappearing dividends: changing firm characteristics or lower propensity to Pay? *Journal of financial economics* 60, 3–44.
- Feng, X., X. He and J. Hu, 2011, Wild bootstrap for quantile regression, *Biometrika* 98, 995–999.
- Galvao, A.F., 2009, Unit root quantile autoregression testing using covariates, *Journal of econometrics* 152, 165–178.
- Garrett, I. and R. Priestley, 2000, Dividend behavior and dividend signaling, *Journal of financial and quantitative analysis* 35, 173–189.
- Granger, C.W.J., 2010, Some thoughts on the development of cointegration, *Journal of econometrics* 158, 3–6
- Hall, P. and S. Sheather, 1988, On the distribution of a studentized quantile, *Journal of royal statistical society, Series B* 50, 381–391.
- Han, C., J. Cho and P.C.B. Phillips, 2011, Infinite density at the median and the typical shape of stock return distributions, *Journal of business & economic statistics* 29, 282–294.
- Hanyunga, D.K. and C. Stephens, 2009, Dividend behavior of U.S. equity REITs, *Journal of property research* 26, 105–123.
- Hurwicz, L., 1950, Least-squares bias in time series, in: T.C. Koopmans (Ed.), *Inference in dynamic economic models*, Wiley, New York, 365–383.
- Kim, T. and H. White, 2003, Estimation, inference, and specification testing for possibly misspecified quantile regression, in: T. Fomby and R. Hill (Eds.), *Maximum likelihood estimation of misspecified models: twenty years later*, Vol. 17. Elsevier, New York, 107–132.

- Knight, K., 1998, Limiting distributions for L_1 regression estimators under general conditions, *Annals of statistics* 26, 755–770.
- Koenker, R. and G. Bassett, 1978, Regression quantiles, *Econometrica* 46, 33–50.
- Koenker, R. and K. Hallock, 2001, Quantile Regression, *Journal of Economic Perspectives* 15, 143–156.
- Koenker, R. and Z. Xiao, 2002, Inference on the quantile regression process, *Econometrica* 70, 1583–1612.
- Koenker, R. and Z. Xiao, 2004, Unit root quantile autoregression inference, *Journal of the American statistical association* 99, 775–787.
- Koenker, R. and Z. Xiao, 2006, Quantile autoregression, *Journal of the American statistical association* 101, 980–990.
- Koenker, R. and Q. Zhao, 1996, Conditional quantile estimation and inference for ARCH models, *Econometric theory* 12, 793–813.
- Koijen, R., J.C. Rodriguez, and A. Sbuely, 2009, Momentum and mean reversion in strategic asset allocation, *Management science* 55, 1199–1213.
- Leary, M.T. and R. Michaely, 2011, Determinants of dividend smoothing: empirical evidence, *Review of financial studies* 24, 3197–3249.
- Lee, C.C. and J.H. Zeng, 2011, Revisiting the relationship between spot and futures oil prices: evidence from quantile cointegrating regression, *Energy economics* 33, 924–935.
- Lee, J.H., 2014, Predictive quantile regression with persistent covariates: IVX-QR approach, unpublished manuscript, University of Washington.
- Lintner, J., 1956, Distribution of incomes of corporations among dividends, retained earnings, and taxes, *American economic review* 46, 97–113.
- Marsh, T. and R. Merton, 1987, Dividend behaviour for the aggregate stock market, *Journal of business* 60, 1–40.
- Martins, P.S. and P.T. Pereira, 2004, Does education reduce wage inequality? quantile regression evidence from 16 countries, *Labour economics* 11, 355–71.

- Maynard, A., K. Shimotsu and Y. Wang, 2011, Inference in predictive quantile regressions, unpublished manuscript, University of Guelph.
- Pesaran, M.H. and Y. Shin, 1998, An autoregressive distributed lag Modelling approach to cointegration analysis, in: S. Strøm, (Ed.), *Econometrics and economic theory in the twentieth century: the Ragnar Frisch centennial symposium*, Cambridge university press, Cambridge, UK, 371–413.
- Pesaran M.H., Y. Shin, and R.J. Smith, 2001, Bounds testing approaches to the analysis of level relationships, *Journal of applied econometrics* 16, 289–326.
- Phillips, P.C.B., 1991, A shortcut to LAD estimator asymptotics, *Econometric theory* 7, 450–463.
- Phillips, P.C.B. and B. Hansen, 1990, Statistical inference in instrumental variables regression with $I(1)$ processes, *Review of economic studies* 57, 99–125.
- Phillips, P.C.B. and V. Solo, 1992, Asymptotics for linear processes, *Annals of statistics* 20, 971–1001.
- Rangvid, J., M. Schmeling and A. Schrimp, 2012, Dividend predictability around the world, *Journal of financial and quantitative analysis* (forthcoming).
- Saikkonen, P., 1991, Asymptotically efficient estimation of cointegration regression, *Econometric theory* 7, 1–21.
- Scott, D.J., 1973, Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach, *Advances in applied probability* 5, 119–137.
- Skinner, D., 2008, The evolving relation between earnings, dividends, and stock repurchases, *Journal of financial economics* 87, 582–609.
- Shiller, R., 2005, *Irrational exuberance*, Princeton university press.
- Shin, Y., B. Yu and M. Greenwood-Nimmo, 2104, Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework, *Festschrift in honor of Peter Schmidt*, Springer, New York.
- Tsong, C.C. and C.F. Lee, 2013, Quantile cointegration analysis of the Fisher hypothesis, *Journal of macroeconomics* 35, 186–198.
- Wang, Y., 2012, *Three essays on time series quantile regression*, PhD Dissertation, Queen’s University.
- Xiao, Z., 2009, Quantile cointegrating regression, *Journal of econometrics* 150, 248–260.

Distributions	Tests \ Sample Size	50	200	400	600	800	1,000	2,000
$U_t \sim N(0, 1)$	$\mathcal{W}_n^{(1)}(\mathbf{B})$	12.24	7.26	6.50	5.82	6.02	5.66	5.20
	$\mathcal{W}_n^{(2)}(\mathbf{B})$	12.66	7.22	6.36	5.88	5.72	5.60	5.46
	$\mathcal{W}_n^{(3)}(\mathbf{B})$	18.96	8.58	6.92	6.22	6.46	5.78	6.20
	$\mathcal{W}_n^{(4)}(\mathbf{B})$	18.54	9.72	7.18	6.20	6.32	5.84	5.42
$U_t \sim t_5$	$\mathcal{W}_n^{(1)}(\mathbf{B})$	14.88	8.06	6.72	6.30	6.42	5.14	5.62
	$\mathcal{W}_n^{(2)}(\mathbf{B})$	14.58	8.10	5.84	6.16	5.00	5.66	5.40
	$\mathcal{W}_n^{(3)}(\mathbf{B})$	20.72	9.26	6.86	6.70	6.20	5.88	5.50
	$\mathcal{W}_n^{(4)}(\mathbf{B})$	22.38	9.70	7.58	6.82	6.20	6.00	5.90

Table 1: EMPIRICAL LEVELS OF $\mathcal{W}_n^{(1)}(\mathbf{B})$, $\mathcal{W}_n^{(2)}(\mathbf{B})$, $\mathcal{W}_n^{(3)}(\mathbf{B})$, AND $\mathcal{W}_n^{(4)}(\mathbf{B})$ (LEVEL OF SIGNIFICANCE: 5%). Notes: (i) Number of iterations: 5,000. (ii) DGP: $Y_t = \alpha_* + \phi_* Y_{t-1} + \theta_{0*} X_t + \theta_{1*} X_{t-1} + U_t$, $X_t := X_{t-1} + W_t$, $W_t := \rho_* R_{t-1} + (1 - \rho_*^2) R_t$, and $R_t \sim \text{IID } N(0, 1)$. (iii) Model: $Y_t = \alpha_*(\tau) + \mathbf{W}'_t \boldsymbol{\delta}_*(\tau) + \mathbf{X}'_t \boldsymbol{\gamma}_*(\tau) + \phi_*(\tau) Y_{t-1} + U_t(\tau)$. (iv) The null hypothesis: $\mathbf{s}_1 = \mathbf{s}_2 = \mathbf{s}_3 = 0$, and $\mathbf{s}_4 = [0, 0]'$ for $\mathcal{W}_n^{(1)}(\mathbf{B})$, $\mathcal{W}_n^{(2)}(\mathbf{B})$, $\mathcal{W}_n^{(3)}(\mathbf{B})$, and $\mathcal{W}_n^{(4)}(\mathbf{B})$, respectively. (v) $\alpha_* = 1$, $\phi_* = 0.25$, $\theta_{0*} = 2$, $\theta_{1*} = 3$, and $\rho_* = 0.5$.

Distributions	Tests \ Sample Size	50	200	400	600	800	1,000	2,000
$U_t \sim N(0, 1)$	$\mathcal{W}_n^{(1)}(\Phi)$	9.80	6.68	5.88	6.08	4.98	4.82	5.40
	$\mathcal{W}_n^{(2)}(\Phi)$	10.66	6.44	5.74	5.88	5.26	5.44	5.14
	$\mathcal{W}_n^{(3)}(\Phi)$	15.04	8.24	6.08	6.74	5.38	5.92	5.84
	$\mathcal{W}_n^{(4)}(\Phi)$	14.90	8.40	6.28	6.96	5.54	5.98	5.76
$U_t \sim t_5$	$\mathcal{W}_n^{(1)}(\Phi)$	11.84	7.56	6.76	6.20	5.40	5.74	4.96
	$\mathcal{W}_n^{(2)}(\Phi)$	11.90	8.12	6.40	6.44	6.16	5.84	5.54
	$\mathcal{W}_n^{(3)}(\Phi)$	18.30	8.98	7.40	7.38	6.32	6.26	5.18
	$\mathcal{W}_n^{(4)}(\Phi)$	18.62	9.64	7.02	7.64	6.28	6.40	5.46

Table 2: EMPIRICAL LEVELS OF $\mathcal{W}_n^{(1)}(\Phi)$, $\mathcal{W}_n^{(2)}(\Phi)$, $\mathcal{W}_n^{(3)}(\Phi)$, AND $\mathcal{W}_n^{(4)}(\Phi)$ (LEVEL OF SIGNIFICANCE: 5%). Notes: Refer to Table 1 for the simulation environments.

Distributions	Tests \ Sample Size	50	200	400	600	800	1,000	2,000
$U_t \sim N(0, 1)$	$\mathcal{W}_n^{(1)}(\Gamma)$	8.08	6.64	5.90	6.00	5.04	4.78	5.42
	$\mathcal{W}_n^{(2)}(\Gamma)$	8.74	6.04	5.58	6.04	5.02	5.50	5.26
	$\mathcal{W}_n^{(3)}(\Gamma)$	13.54	7.84	5.84	6.60	5.22	5.86	5.92
	$\mathcal{W}_n^{(4)}(\Gamma)$	12.70	7.68	6.14	6.82	5.46	5.84	5.60
$U_t \sim t_5$	$\mathcal{W}_n^{(1)}(\Gamma)$	10.10	7.28	6.60	6.28	5.48	5.74	4.92
	$\mathcal{W}_n^{(2)}(\Gamma)$	10.36	7.66	6.08	6.32	6.08	5.78	5.52
	$\mathcal{W}_n^{(3)}(\Gamma)$	16.58	8.38	7.06	7.28	6.14	6.30	5.08
	$\mathcal{W}_n^{(4)}(\Gamma)$	15.68	8.84	6.96	7.64	6.14	6.24	5.66

Table 3: EMPIRICAL LEVELS OF $\mathcal{W}_n^{(1)}(\Gamma)$, $\mathcal{W}_n^{(2)}(\Gamma)$, $\mathcal{W}_n^{(3)}(\Gamma)$, AND $\mathcal{W}_n^{(4)}(\Gamma)$ (LEVEL OF SIGNIFICANCE; 5%). Notes: Refer to Table 1 for the simulation environments.

Distributions	Tests \ Sample Size	50	100	200	300	400	500
$U_t \sim N(0, 1)$	$\mathcal{W}_n^{(1)}(\mathbf{B})$	41.75	74.90	97.80	99.85	99.95	100.0
	$\mathcal{W}_n^{(2)}(\mathbf{B})$	42.50	75.00	98.00	99.75	99.90	100.0
	$\mathcal{W}_n^{(3)}(\mathbf{B})$	37.00	64.15	93.30	98.95	99.85	100.0
	$\mathcal{W}_n^{(4)}(\mathbf{B})$	64.20	90.20	99.70	100.0	100.0	100.0
$U_t \sim t_5$	$\mathcal{W}_n^{(1)}(\mathbf{B})$	42.62	72.68	95.70	99.56	99.98	100.0
	$\mathcal{W}_n^{(2)}(\mathbf{B})$	41.36	70.96	95.66	99.70	100.0	100.0
	$\mathcal{W}_n^{(3)}(\mathbf{B})$	37.86	59.10	88.70	97.84	99.60	100.0
	$\mathcal{W}_n^{(4)}(\mathbf{B})$	63.12	87.96	99.40	100.0	100.0	100.0

Table 4: EMPIRICAL POWERS OF $\mathcal{W}_n^{(1)}(\mathbf{B})$, $\mathcal{W}_n^{(2)}(\mathbf{B})$, $\mathcal{W}_n^{(3)}(\mathbf{B})$, AND $\mathcal{W}_n^{(4)}(\mathbf{B})$ (LEVEL OF SIGNIFICANCE: 5%). Notes: Refer to Table 1 for the simulation environments.

Distributions	Tests \ Sample Size	50	100	200	300	400	500
$U_t \sim N(0, 1)$	$\mathcal{W}_n^{(1)}(\Phi)$	78.66	96.42	99.98	100.0	100.0	100.0
	$\mathcal{W}_n^{(2)}(\Phi)$	79.18	96.12	100.0	100.0	100.0	100.0
	$\mathcal{W}_n^{(3)}(\Phi)$	65.22	85.50	98.52	99.88	100.0	100.0
	$\mathcal{W}_n^{(4)}(\Phi)$	97.56	99.92	100.0	100.0	100.0	100.0
$U_t \sim t_5$	$\mathcal{W}_n^{(1)}(\Phi)$	74.58	93.92	99.90	100.0	100.0	100.0
	$\mathcal{W}_n^{(2)}(\Phi)$	73.22	93.72	99.82	99.98	100.0	100.0
	$\mathcal{W}_n^{(3)}(\Phi)$	61.32	80.34	96.18	99.36	99.90	99.98
	$\mathcal{W}_n^{(4)}(\Phi)$	94.40	99.68	100.0	100.0	100.0	100.0

Table 5: EMPIRICAL POWERS OF $\mathcal{W}_n^{(1)}(\Phi)$, $\mathcal{W}_n^{(2)}(\Phi)$, $\mathcal{W}_n^{(3)}(\Phi)$, AND $\mathcal{W}_n^{(4)}(\Phi)$ (LEVEL OF SIGNIFICANCE: 5%). Notes: Refer to Table 1 for the simulation environments.

Distributions	Tests \ Sample Size	200	400	600	800	1,200	1,600	2,000
$U_t \sim N(0, 1)$	$\mathcal{W}_n^{(1)}(\Gamma)$	15.70	22.10	29.82	37.54	49.78	62.20	72.94
	$\mathcal{W}_n^{(2)}(\Gamma)$	15.12	21.80	29.72	36.28	49.96	61.80	71.04
	$\mathcal{W}_n^{(3)}(\Gamma)$	13.68	16.64	19.60	24.84	32.80	41.86	50.22
	$\mathcal{W}_n^{(4)}(\Gamma)$	24.22	36.48	49.72	61.30	77.96	89.12	94.84
$U_t \sim t_5$	$\mathcal{W}_n^{(1)}(\Gamma)$	14.48	19.56	26.96	33.38	43.38	56.16	64.40
	$\mathcal{W}_n^{(2)}(\Gamma)$	14.62	20.36	26.28	32.42	44.76	56.16	64.30
	$\mathcal{W}_n^{(3)}(\Gamma)$	13.76	15.14	18.32	22.02	27.78	36.12	43.12
	$\mathcal{W}_n^{(4)}(\Gamma)$	22.36	31.68	42.72	53.20	68.62	81.86	88.52

Table 6: EMPIRICAL POWERS OF $\mathcal{W}_n^{(1)}(\Gamma)$, $\mathcal{W}_n^{(2)}(\Gamma)$, $\mathcal{W}_n^{(3)}(\Gamma)$, AND $\mathcal{W}_n^{(4)}(\Gamma)$ (LEVEL OF SIGNIFICANCE: 5%). Notes: Refer to Table 1 for the simulation environments.

		$\widetilde{\mathcal{W}}_n^{(1)}(\mathbf{B})$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	13.60	7.68	6.68	5.68	6.26	5.92	5.28
	0.10	13.46	7.66	7.10	6.48	5.90	5.74	5.12
	0.20	12.98	7.70	6.76	5.74	5.78	5.92	5.28
	0.30	14.32	7.52	7.32	6.60	6.22	5.80	5.40
	0.40	13.08	8.28	7.24	6.00	6.48	6.06	5.90
		$\widetilde{\mathcal{W}}_n^{(2)}(\mathbf{B})$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	13.84	7.16	6.44	5.44	6.06	5.44	5.92
	0.10	13.48	8.80	7.04	6.30	5.96	5.34	4.72
	0.20	12.32	7.88	6.54	6.24	5.92	5.82	5.20
	0.30	13.64	7.48	7.04	6.40	5.32	5.62	5.46
	0.40	12.14	6.98	6.64	6.74	5.18	5.80	5.58
		$\widetilde{\mathcal{W}}_n^{(3)}(\mathbf{B})$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	21.44	9.00	7.56	6.34	6.92	6.12	5.62
	0.10	20.38	9.36	8.32	7.18	6.58	6.60	4.98
	0.20	19.96	9.90	7.52	6.68	5.88	6.76	5.22
	0.30	21.50	9.86	8.28	7.24	6.24	6.60	5.80
	0.40	20.36	10.04	8.52	7.44	6.82	6.52	6.58
		$\widetilde{\mathcal{W}}_n^{(4)}(\mathbf{B})$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	21.64	9.42	7.68	6.10	7.10	6.00	5.98
	0.10	20.66	10.00	7.86	7.20	6.36	6.28	5.32
	0.20	20.06	9.80	7.28	6.80	6.64	6.74	4.96
	0.30	21.84	9.36	8.66	7.24	6.56	6.24	5.28
	0.40	19.60	9.84	7.92	7.42	6.54	6.26	6.00

Table 7: EMPIRICAL LEVELS OF $\widetilde{\mathcal{W}}_n^{(1)}(\mathbf{B})$, $\widetilde{\mathcal{W}}_n^{(2)}(\mathbf{B})$, $\widetilde{\mathcal{W}}_n^{(3)}(\mathbf{B})$, AND $\widetilde{\mathcal{W}}_n^{(4)}(\mathbf{B})$ (LEVEL OF SIGNIFICANCE: 5%). Notes: (i) Number of iterations: 5,000. (ii) DGP: $Y_t = \alpha_* + \phi_* Y_{t-1} + \theta_{0*} X_t + \theta_{1*} X_{t-1} + U_t$, $X_t := X_{t-1} + W_t$, $U_t := \sigma_* R_{t-1}^{(1)} + (1 - \sigma_*^2) R_t^{(1)}$, $W_t := \rho_* R_{t-1}^{(2)} + (1 - \rho_*^2) R_t^{(2)}$, and $(R_t^{(1)}, R_t^{(2)})' \sim \text{IID } N(\mathbf{0}, \mathbf{I}_2)$ with $\alpha_* = 1$, $\phi_* = 0.25$, $\theta_{0*} = 2$, $\theta_{1*} = 3$, and $\rho_* = 0.5$. (iii) Model: $Y_t = \alpha_*(\tau) + \sum_{j=0}^{q-1} \mathbf{W}'_{t-j} \boldsymbol{\delta}_{j*}(\tau) + \mathbf{X}'_t \boldsymbol{\gamma}_*(\tau) + \sum_{j=1}^p \phi_{j*}(\tau) Y_{t-j} + U_t(\tau)$. (iv) The null hypothesis: $\mathbf{s}_1 = \mathbf{s}_2 = \mathbf{s}_3 = \mathbf{0}$, and $\mathbf{s}_4 = [0, 0]'$ for $\widetilde{\mathcal{W}}_n^{(1)}(\mathbf{B})$, $\widetilde{\mathcal{W}}_n^{(2)}(\mathbf{B})$, $\widetilde{\mathcal{W}}_n^{(3)}(\mathbf{B})$, and $\widetilde{\mathcal{W}}_n^{(4)}(\mathbf{B})$, respectively. (v) BIC is applied to determine the QARDL orders p and q .

		$\widetilde{\mathcal{W}}_n^{(1)}(\Phi)$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	12.58	7.48	6.70	5.54	5.42	5.54	5.42
	0.10	12.66	8.14	6.84	6.22	5.78	5.68	5.32
	0.20	12.00	7.26	6.74	6.32	5.50	5.26	5.14
	0.30	10.88	7.32	6.62	6.24	5.60	4.68	5.06
	0.40	10.00	7.04	6.22	5.94	6.38	5.74	5.34
		$\widetilde{\mathcal{W}}_n^{(2)}(\Phi)$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	13.66	7.56	6.20	5.78	5.46	6.22	6.36
	0.10	11.42	7.62	6.74	6.06	5.74	5.42	5.08
	0.20	12.00	7.66	6.00	5.92	6.10	5.50	4.98
	0.30	11.20	7.20	6.84	6.12	5.62	5.78	4.86
	0.40	10.38	7.08	6.32	5.68	6.08	5.60	5.18
		$\widetilde{\mathcal{W}}_n^{(3)}(\Phi)$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	20.38	9.46	7.32	6.46	6.02	6.38	6.32
	0.10	19.68	10.24	7.18	7.00	6.30	6.20	5.04
	0.20	19.46	9.56	7.16	6.44	6.26	6.30	4.94
	0.30	16.40	8.78	7.08	7.26	6.42	5.44	5.28
	0.40	16.90	8.94	7.14	6.34	6.48	6.14	6.02
		$\widetilde{\mathcal{W}}_n^{(4)}(\Phi)$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	20.78	9.20	7.58	5.96	6.06	6.20	6.44
	0.10	19.16	9.72	7.42	6.64	6.38	5.70	5.32
	0.20	18.78	9.40	7.24	6.84	5.90	5.76	4.92
	0.30	16.68	8.84	7.32	7.06	6.14	5.42	5.44
	0.40	15.72	8.96	7.20	6.36	6.82	6.06	5.60

Table 8: EMPIRICAL LEVELS OF $\widetilde{\mathcal{W}}_n^{(1)}(\Phi)$, $\widetilde{\mathcal{W}}_n^{(2)}(\Phi)$, $\widetilde{\mathcal{W}}_n^{(3)}(\Phi)$, AND $\widetilde{\mathcal{W}}_n^{(4)}(\Phi)$ (LEVEL OF SIGNIFICANCE: 5%). Notes: Refer to Table 7 for the simulation environments.

		$\widetilde{\mathcal{W}}_n^{(1)}(\Gamma)$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	12.26	7.32	6.70	5.44	5.42	5.66	5.44
	0.10	11.78	8.02	6.70	6.16	5.82	5.60	5.26
	0.20	11.46	7.04	6.74	6.38	5.50	5.20	5.18
	0.30	10.36	7.18	6.58	6.10	5.48	4.68	5.10
	0.40	9.74	6.68	6.10	5.92	6.46	5.68	5.32
		$\widetilde{\mathcal{W}}_n^{(2)}(\Gamma)$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	13.06	7.36	6.14	5.78	5.42	6.26	6.36
	0.10	11.00	7.40	6.78	6.08	5.82	5.44	5.10
	0.20	11.52	7.48	6.00	5.98	6.06	5.58	5.04
	0.30	10.62	7.14	6.78	6.12	5.56	5.74	4.86
	0.40	9.90	7.04	6.30	5.70	6.04	5.56	5.18
		$\widetilde{\mathcal{W}}_n^{(3)}(\Gamma)$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	19.66	9.22	7.32	6.32	6.08	6.40	6.42
	0.10	19.28	10.16	7.12	6.96	6.30	6.14	5.04
	0.20	18.94	9.56	7.04	6.42	6.24	6.18	4.90
	0.30	15.74	8.58	7.02	7.08	6.32	5.48	5.34
	0.40	16.36	8.88	7.24	6.32	6.42	6.26	6.02
		$\widetilde{\mathcal{W}}_n^{(4)}(\Gamma)$						
$\sigma_* \setminus$ Sample Size		50	200	400	600	800	1,000	2,000
	0.00	20.18	9.12	7.56	5.98	6.02	6.10	6.44
	0.10	18.42	9.58	7.24	6.68	6.22	5.54	5.34
	0.20	18.16	9.08	7.12	6.72	5.84	5.80	4.82
	0.30	15.42	8.76	7.10	6.94	6.08	5.38	5.40
	0.40	14.60	8.88	7.00	6.34	6.68	6.02	5.42

Table 9: EMPIRICAL LEVELS OF $\widetilde{\mathcal{W}}_n^{(1)}(\Gamma)$, $\widetilde{\mathcal{W}}_n^{(2)}(\Gamma)$, $\widetilde{\mathcal{W}}_n^{(3)}(\Gamma)$, AND $\widetilde{\mathcal{W}}_n^{(4)}(\Gamma)$ (LEVEL OF SIGNIFICANCE: 5%). Notes: Refer to Table 7 for the simulation environments.

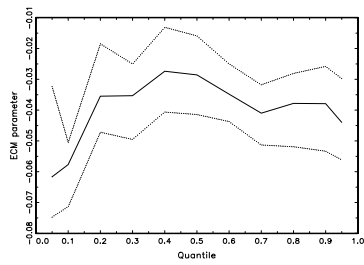
Empirical Level							
Tests	$\sigma_* \setminus$ Sample Size	50	200	400	600	800	1,000
$\mathcal{W}_n(\beta)$	0.00	13.36	6.76	5.86	5.00	5.54	5.56
	0.10	15.80	7.36	6.80	6.52	6.24	5.70
	0.20	16.18	7.48	6.12	4.92	5.00	5.02
	0.30	17.04	7.36	5.10	5.10	5.62	5.18
	0.40	18.02	7.64	6.08	6.32	6.72	5.92
$\mathfrak{W}_n(\beta)$	0.00	17.02	7.84	6.52	6.10	5.68	5.28
	0.10	19.22	9.96	7.50	7.42	6.96	7.44
	0.20	22.04	10.76	9.56	9.24	8.50	8.54
	0.30	24.12	12.70	11.30	10.78	10.50	10.36
	0.40	26.98	14.60	12.70	11.78	11.30	11.16
Empirical Power							
Tests	$\sigma_* \setminus$ Sample Size	50	100	200	300	400	500
$\mathcal{W}_n(\beta)$	0.00	54.22	85.34	99.22	99.94	100.0	100.0
	0.10	50.86	83.10	80.84	99.94	100.0	100.0
	0.20	49.98	80.84	98.56	99.94	100.0	100.0
	0.30	48.30	78.02	97.08	99.82	100.0	100.0
	0.40	46.62	74.32	96.78	99.76	100.0	100.0
$\mathfrak{W}_n(\beta)$	0.00	53.12	85.44	99.10	99.96	100.0	100.0
	0.10	52.58	85.36	99.04	100.0	100.0	100.0
	0.20	55.68	84.60	99.10	99.98	100.0	100.0
	0.30	56.60	84.40	99.00	99.96	100.0	100.0
	0.40	56.30	85.58	99.18	99.94	100.0	100.0

Table 10: EMPIRICAL REJECTION RATES OF $\mathcal{W}_n(\beta)$ AND $\mathfrak{W}_n(\beta)$ (LEVEL OF SIGNIFICANCE: 5%). Notes: $\mathcal{W}_n(\beta)$ and $\mathfrak{W}_n(\beta)$ indicate the Wald test of this paper and the augmented model-based Wald test in Xiao (2009), respectively. The current simulation environments are identical to those in Table 7 with the only difference that ϕ_* is modified to 0 from 0.25. The empirical level is obtained by testing $\beta(0.5) = (\theta_{0*} + \theta_{1*})/(1 - \phi_*)$, and the empirical power is obtained by testing $\beta(0.5) = (\theta_{0*} + \theta_{1*})/(1 - \phi_*) + 0.1$.

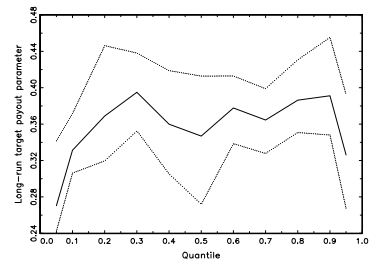
(a) OLS Estimation Results				
ζ_*	β_*	λ_*	δ_*	
-0.04	0.36	0.48	0.01	
(0.005)	(0.013)	(0.039)	0.007	

(b) Quantile Estimation Results				
Quantile index (τ)	$\zeta_*(\tau)$	$\beta_*(\tau)$	$\lambda_*(\tau)$	$\delta_*(\tau)$
0.05	-0.06	0.27	0.58	0.02
	(0.014)	(0.030)	(0.043)	(0.015)
0.10	-0.06	0.33	0.55	0.02
	(0.006)	(0.020)	(0.051)	(0.012)
0.20	-0.04	0.37	0.55	0.02
	(0.009)	(0.050)	(0.057)	(0.007)
0.30	-0.04	0.40	0.53	0.02
	(0.007)	(0.027)	(0.053)	(0.010)
0.40	-0.03	0.36	0.48	0.02
	(0.008)	(0.041)	(0.059)	(0.012)
0.50	-0.03	0.35	0.47	0.03
	(0.008)	(0.043)	(0.066)	(0.013)
0.60	-0.03	0.38	0.50	0.02
	(0.006)	(0.023)	(0.067)	(0.013)
0.70	-0.04	0.36	0.45	0.02
	(0.006)	(0.022)	(0.065)	(0.012)
0.80	-0.04	0.39	0.37	0.01
	(0.008)	(0.024)	(0.053)	(0.008)
0.90	-0.04	0.39	0.30	-0.00
	(0.009)	(0.033)	(0.063)	(0.008)
0.95	-0.04	0.33	0.44	-0.01
	(0.008)	(0.039)	(0.050)	(0.009)

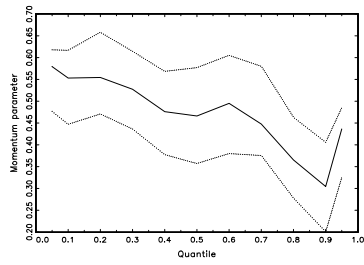
Table 11: OLS AND QUANTILE ESTIMATION RESULTS BASED ON THE WHOLE SAMPLE. Notes: (i) Standard errors are in parentheses, and those for the long-run coefficient are calculated via the delta method. (ii) OLS estimation results are based on the following model: $\Delta D_t = \alpha_* + \zeta_* D_{t-1} + \gamma_* E_{t-1} + \lambda_* \Delta D_{t-1} + \delta_* \Delta E_t + U_t = \alpha_* + \zeta_* (D_{t-1} - \beta_* E_{t-1}) + \lambda_* \Delta D_{t-1} + \delta_* \Delta E_t + U_t$. (iii) Quantile estimation results are based on the following model: $\Delta D_t = \alpha_*(\tau) + \zeta_*(\tau) D_{t-1} + \gamma_*(\tau) E_{t-1} + \lambda_*(\tau) \Delta D_{t-1} + \delta_*(\tau) \Delta E_t + U_t(\tau) = \alpha_*(\tau) + \zeta_*(\tau) (D_{t-1} - \beta_*(\tau) E_{t-1}) + \lambda_*(\tau) \Delta D_{t-1} + \delta_*(\tau) \Delta E_t + U_t(\tau)$.



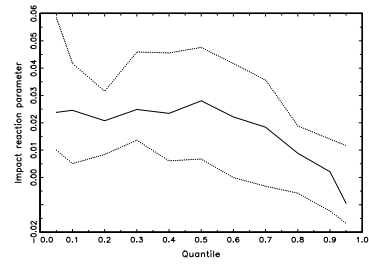
ECM Parameter ($\zeta_*(\tau)$)



Long-Run Target Payout Parameter ($\beta_*(\tau)$)



Momentum Parameter ($\lambda_*(\tau)$)



Impact Reaction Parameter ($\delta_*(\tau)$)

Figure 1: PARAMETER ESTIMATES USING THE WHOLE SAMPLE. These are estimated parameters (the middle solid line) using all available observations for different quantile levels: 0.05, 0.10, ... 0.95, with 90% confidence intervals (the outer dotted lines).

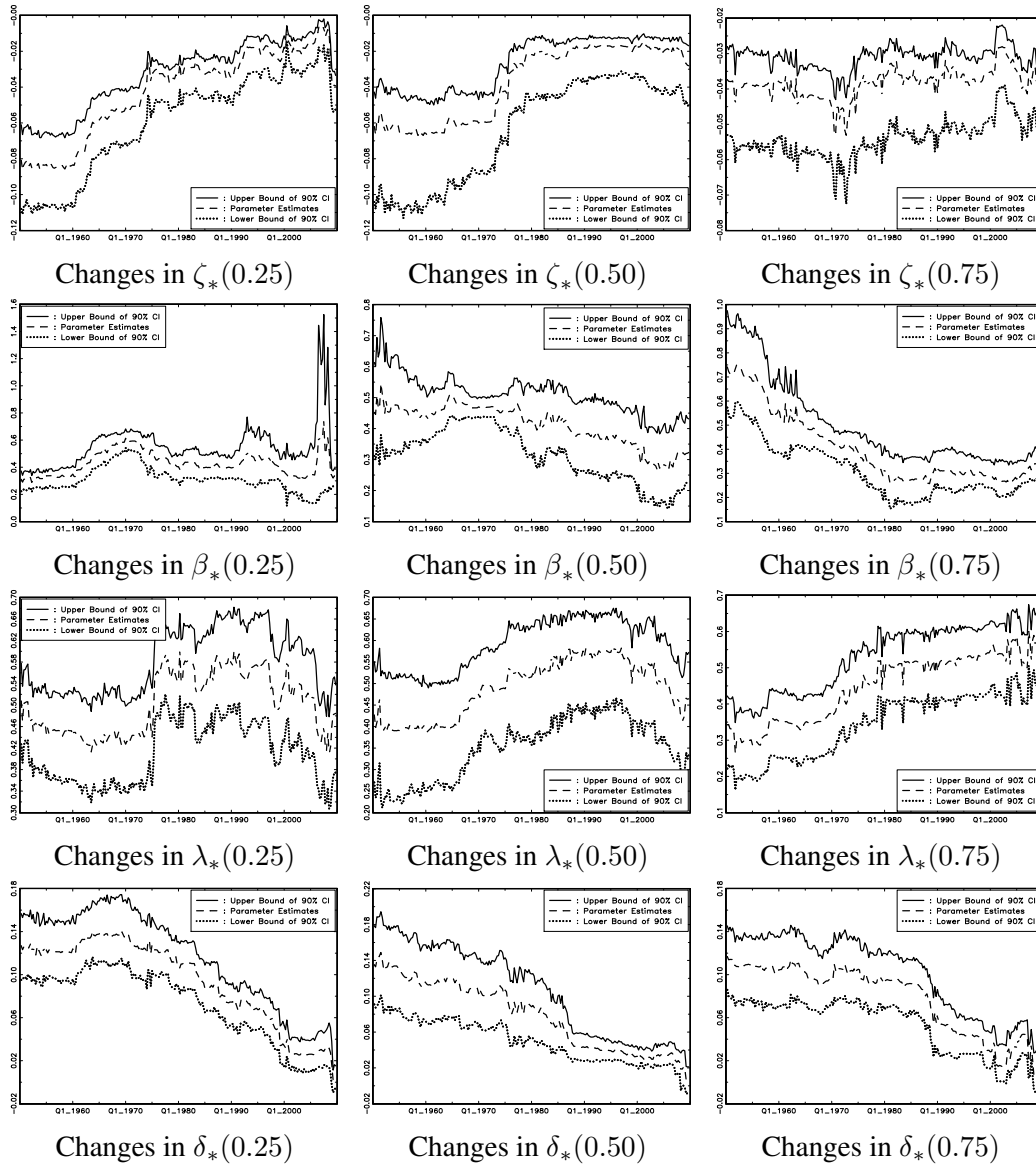


Figure 2: PARAMETER ESTIMATES $\zeta_*(\tau)$, $\beta_*(\tau)$, $\lambda_*(\tau)$, AND $\delta_*(\tau)$ USING THE ROLLING WINDOW METHOD. These are estimated parameters using the rolling window method, and each window has 320 observations. Three different quantile levels are considered: 0.25, 0.5, 0.75. The horizontal axis indicates the last date for the corresponding estimation window. For example, the first estimation window uses observations from 1871Q1 to 1950Q4 so that the first date on the horizontal axis is 1950Q4. The number of the out-of-sample observations is 239.

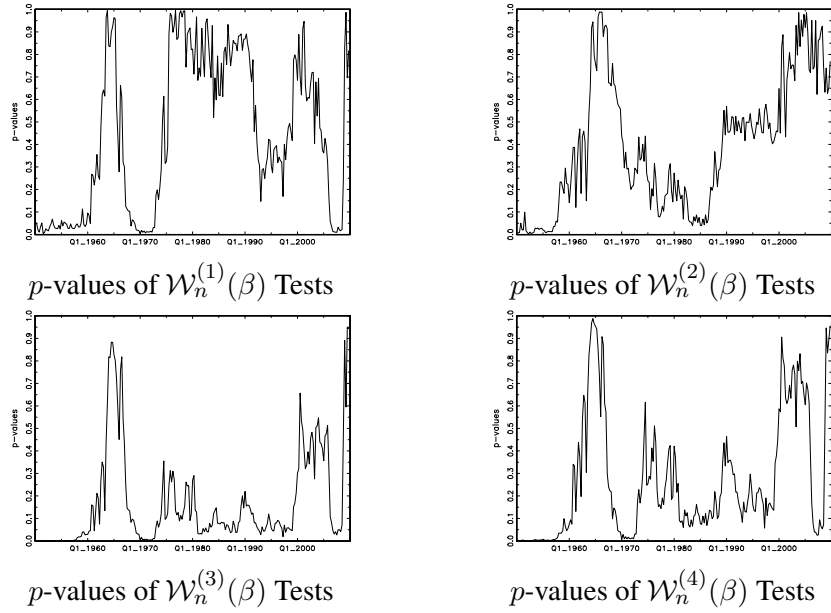


Figure 3: p -VALUES OF $\mathcal{W}_n(\beta)$ TEST STATISTICS. The figures show the estimated p -values of the Wald tests, where $\mathcal{W}_n^{(1)}(\beta)$ tests $\beta_*(0.25) = \beta_*(0.5)$; $\mathcal{W}_n^{(2)}(\beta)$ tests $\beta_*(0.5) = \beta_*(0.75)$; $\mathcal{W}_n^{(3)}(\beta)$ tests $\beta_*(0.25) = \beta_*(0.75)$; and $\mathcal{W}_n^{(4)}(\beta)$ tests $\beta_*(0.25) = \beta_*(0.5) = \beta_*(0.75)$. The horizontal axis indicates the last date for the corresponding estimation window. The number of the out-of-sample observations is 239.

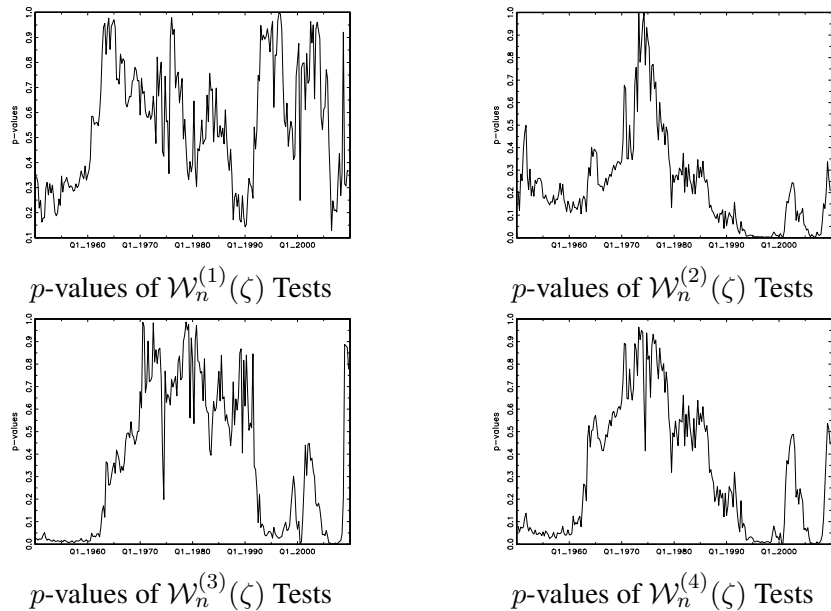


Figure 4: p -VALUES OF $\mathcal{W}_n(\zeta)$ TEST STATISTICS. The figures show the estimated p -values of the Wald tests, where $\mathcal{W}_n^{(1)}(\zeta)$ tests $\zeta_*(0.25) = \zeta_*(0.5)$; $\mathcal{W}_n^{(2)}(\zeta)$ tests $\zeta_*(0.5) = \zeta_*(0.75)$; $\mathcal{W}_n^{(3)}(\zeta)$ tests $\zeta_*(0.25) = \zeta_*(0.75)$; and $\mathcal{W}_n^{(4)}(\zeta)$ tests $\zeta_*(0.25) = \zeta_*(0.5) = \zeta_*(0.75)$. See notes to Figure 3.

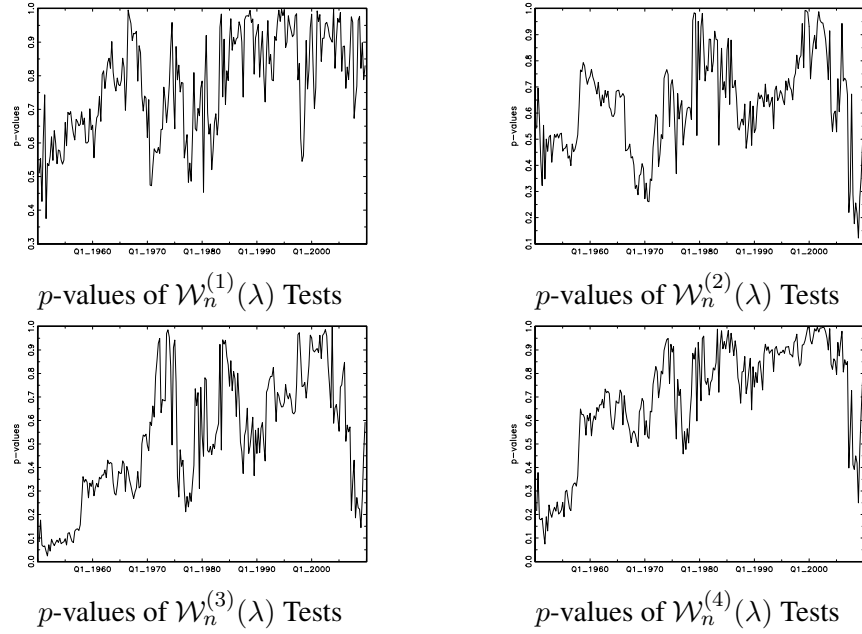


Figure 5: p -VALUES OF $\mathcal{W}_n(\lambda)$ TEST STATISTICS. The figures show the estimated p -values of the Wald tests, where $\mathcal{W}_n^{(1)}(\lambda)$ tests $\lambda_*(0.25) = \lambda_*(0.5)$; $\mathcal{W}_n^{(2)}(\lambda)$ tests $\lambda_*(0.5) = \lambda_*(0.75)$; $\mathcal{W}_n^{(3)}(\lambda)$ tests $\lambda_*(0.25) = \lambda_*(0.75)$; and $\mathcal{W}_n^{(4)}(\lambda)$ tests $\lambda_*(0.25) = \lambda_*(0.5) = \lambda_*(0.75)$. See notes to Figure 3.

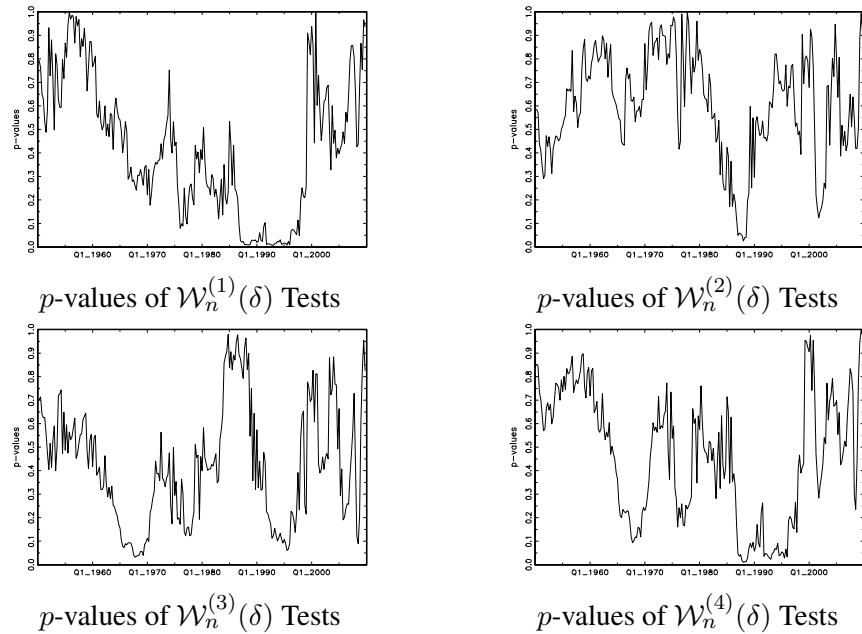


Figure 6: p -VALUES OF $\mathcal{W}_n(\delta)$ TEST STATISTICS. The figures show the estimated p -values of the Wald tests, $\mathcal{W}_n^{(2)}(\delta)$ tests $\delta_*(0.5) = \delta_*(0.75)$; $\mathcal{W}_n^{(3)}(\delta)$ tests $\delta_*(0.25) = \delta_*(0.75)$; and $\mathcal{W}_n^{(4)}(\delta)$ tests $\delta_*(0.25) = \delta_*(0.5) = \delta_*(0.75)$. See notes to Figure 3.