Journal of Econometrics 234 (2023) 227–250

Contents lists available at ScienceDirect

Journal of Econometrics

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

A new robust inference for predictive quantile regression

Zongwu Cai [a], Haiqiang Chen [b,*], Xiaosai Liao [c]

[a] Department of Economics, University of Kansas, Lawrence, KS 66045, USA
[b] Wang Yanan Institute for Studies in Economics and MOE Key Laboratory of Econometrics, Xiamen University, Xiamen, Fujian 361005, China
[c] Institute of Chinese Financial Studies and Collaborative Innovation Center of Financial Security, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China

ARTICLE INFO

Article history:
Received 16 June 2020
Received in revised form 20 October 2021
Accepted 22 October 2021
Available online 20 January 2022

JEL classification:
C32
C52
C58

Keywords:
Auxiliary regressor
Embedded endogeneity
Highly persistent predictor
Multiple regression
Predictive quantile regression
Robust
Weighted estimator

ABSTRACT

This paper proposes a novel approach to offer a robust inferential theory across all types of persistent regressors in a predictive quantile regression model. We first estimate a quantile regression with an auxiliary regressor, which is generated as a weighted combination of an exogenous random walk process and a bounded transformation of the original regressor. With a similar spirit of rotation in factor analysis, one can then construct a weighted estimator using the estimated coefficients of the original predictor and the auxiliary regressor. Under some mild conditions, it shows that the self-normalized test statistic based on the weighted estimator converges to a standard normal distribution. Our new approach enjoys a good property that it can reach the local power under the optimal rate T with nonstationary predictor and $\sqrt{T}$ for stationary predictor, respectively. More importantly, our approach can be easily used to characterize mixed persistency degrees in multiple regressions. Simulations and empirical studies are provided to demonstrate the effectiveness of the newly proposed approach. The heterogeneous predictability of US stock returns at different quantile levels is reexamined.

1. Introduction

A long-term issue in financial econometrics is to test whether or not an asset return process is predictable by a set of lagged predictors (say, financial ratios or/and macroeconomic variables), but the conclusions based on predictive mean regressions are mixed despite an enormous amount of efforts devoted to this problem in the literature, see, for example, the papers by Ang and Bekaert (2007) and Welch and Goyal (2008). The indefinite conclusions are partially due to the econometric issues caused by those highly persistent regressors where conventional test statistics are invalid with a serious size distortion, which is more serious if the innovations of predictors and return errors are contemporarily correlated, as studied by Campbell and Yogo (2006), Torous et al. (2004), Zhu et al. (2014), Choi et al. (2016), Yang et al. (2019), and among others.[1] The heterogeneous predictability of asset returns could be another reason. For example, a stronger prediction power is usually found in recession periods for stock markets; see Gonzalo and Pitarakis (2012), which

* Corresponding author.
E-mail address: hc335@xmu.edu.cn (H. Chen).

[1] In the framework of mean regressions, several solutions were proposed in literature, such as the Bonferroni's method by Campbell and Yogo (2006), the conditional likelihood method by Jansson and Moreira (2006), the linear projection method by Cai and Wang (2014), the instrumental variable (IVX) approach by Magdalinos and Phillips (2009), Phillips and Magdalinos (2009), Kostakis et al. (2015), Phillips and Lee (2016), Yang et al. (2019), Hosseinkouchack and Demetrescu (2021) and Demetrescu and Rodrigues (2020), the Cauchy type instrumental variable approach by Choi

https://doi.org/10.1016/j.jeconom.2021.10.012
0304-4076/

implies potentially greater predictability at lower quantiles. As mean regressions reflect the average predictability over all quantiles, they may fail to find evidence for the predictability of asset returns at some quantiles, particularly in tails. That has motivated researchers to examine the predictability of asset returns using quantile regressions, which reveal more information about the predicability under the entire underlying conditional distribution; see, for example, the papers by Koenker (2005), Xiao (2009), Lee (2016) and Fan and Lee (2019) for details.

Testing the predictability in a quantile setting is of importance in economics and finance and also of practical attractiveness. First, from economic perspective, empirical evidences have documented that investors' interest in asset returns is beyond their mean and variance. For example, Harvey and Siddique (2000) and Dittmar (2002) found that the higher order moments are helpful to explain cross-sectional variation in US stock returns, whereas Cenesizoglu and Timmermann (2008) concluded that the entire distribution of future stock returns is informative for investment decisions of risk averse investors. Second, from the econometric point of view, quantile regressions are more suitable when the distribution is skewed and/or heavy tailed, which is a stylized fact in financial econometrics, and the quantile regression technique has been applied widely in risk management operations. For example, the Value-at-Risk is defined by the unconditional/conditional quantile and is widely used to measure the tail risk in practice. Finally, predictive quantile regressions avoid the order-imbalance issue, a well known theoretical challenge that arises for mean regressions where the dependent variable commonly behaves as martingale differences, while the regressors, fundamental variables, are highly persistent as argued in Phillips (2015).

Modeling predictive quantiles and examining their predictability with possible nonstationary regressors are not trivial. Some challenging econometric issues in mean regressions causing the failure of traditionally econometric inferences of the predictive regression still exist for predictive quantile regressions. To overcome these difficulties, Maynard et al. (2011) tried to extend the Bonferroni's method to predictive quantile regressions with highly persistent regressors. However, as pointed out by Phillips (2015), the Bonferroni's method is hard to be extended to multivariate cases, which are typical in applied research. To the best of our knowledge, the papers by Lee (2016) and Fan and Lee (2019) were the first to investigate the asymptotic theory for multivariate predictive quantile regressions with both various degrees of persistency and embedded endogeneity. Indeed, Lee (2016) extended the instrumental variable filtering methodology by Magdalinos and Phillips (2009), Phillips and Magdalinos (2009), and Kostakis et al. (2015) for mean regressions to quantile regressions, termed as IVX-QR approach. Further, Lee (2016) obtained the asymptotic distribution of test statistics that are robust to the degree of persistency under the null hypothesis. Recently, Fan and Lee (2019) applied the IVX-QR method in Lee (2016) to the situation with conditionally heteroskedastic errors. The key idea of the IVX-QR approach is to generate a mildly integrated process as the instrument of the persistent and possibly endogenous regressor, by which it succeeds in correcting the size distortion, but sacrifices some convergence rates which may lead to a loss of the power. As pointed out by Lee (2016), the improvement of the size control and the magnitude of the power loss are similar to the two sides of a coin, relying on the choice of the filtering parameters in the generation of the mildly integrated process. Though Lee (2016) provided a practical rule for choosing these parameters, it is still unclear how these tuning parameters affect the test performance exactly.

The main contribution of this paper is to propose a novel and easy-to-implement approach, termed as the double weighted method, to develop a uniform (robust) inferential theory for predictive quantile regressions with highly persistent variables. The newly proposed method is based on a quantile regression with an auxiliary regressor, which is generated as a weighted combination of an exogenous simulated nonstationary process and a bounded transformation of the original regressor. The weight, which plays a role of filtering, is chosen by a data-driven approach, such that the auxiliary regressor enjoys having the same persistency degree with the original predictor asymptotically. In addition, to avoid efficiency loss as well as eliminate the impact of the embedded endogeneity, we construct a weighted estimator using the coefficients of both original regressor and auxiliary regressor, with a similar idea of rotation. Under some mild conditions, it shows that the self-normalized test statistics based on the weighted estimator converge to a standard normal or χ^2-distribution. Comparing to the IVX-QR approach, our method possesses a good property that the weighted estimator reaches the local power under the optimal convergence rate T with nonstationary predictors and $\sqrt{T}$ with stationary predictors, respectively. Meanwhile, our method can be easily generalized to multiple regressors with mixed persistency degrees, and allows for testing of a general linear hypothesis of coefficients, which is seminal in the related econometrics literature. Simulations are conducted to demonstrate the effectiveness of our newly proposed approach. For most cases, our method has better size control and power performance in a finite sample compared with other existing methods.

Indeed, our motivation for this study is to implement the newly proposed approach for re-examining the predictability of US stock market returns using eight popular financial ratios and macroeconomic indicators. For the convenience of comparison, the same data set used by Lee (2016) is taken with the sample period from 1927 to 2005. To see whether there is any change after the 2008 global crisis, the data set is updated to December of 2018. The main empirical findings can be summarized as follows. First, the predictability for the middle quantile levels is weaker than both lower and upper quantiles, which is consistent with the previous findings. Second, in the multivariate prediction quantile regression, many variables lose their prediction power after controlling other variables. Third, after the World War II, we do not find much

et al. (2016), the weighted empirical likelihood approach by Zhu et al. (2014), Liu et al. (2019), and Yang et al. (2021), and the variable addition (VA) or augmented regression or control function approach by Elliott (2011) and Breitung and Demetrescu (2015) and Yang et al. (2021). The reader is referred to the recent survey paper by Liao et al. (2018) for more discussions.

evidence of the prediction power for some well-known financial ratios, such as earnings to price (d/p) ratio, dividend to price (d/p) ratio and book to market (b/m) ratio. However, the macroeconomic indicators, like T-bill rate (tbl), default yield spread (dfy), term spread (tms), show some strong evidence of significant prediction power, especially at lower and upper quantile levels. The detailed analysis of this empirical example is reported in Section 5.

In some way, our paper is tied to the regression with auxiliary variables. Indeed, Toda and Yamamoto (1995), and Dolado and Lütkepohl (1996) first proposed a robust testing strategy irrespective of the persistency type of regressor through a regression with additional (redundant) variables, such that the coefficients to be tested are attached to stationary variables, whereas Bauer and Maynard (2012) considered the VA approach in the context of vector autoregressive processes with unknown persistency. In particular, Breitung and Demetrescu (2015) argued that the traditional VA approaches suffer from a loss of power. Different from Breitung and Demetrescu (2015), our paper particularly constructs the additional regressor in its own way and proposes a new test statistic.

The rest of this paper is organized as follows. Section 2 introduces the model framework and provides the procedures for estimating parameters and constructing the test statistics as well as presents the asymptotic theories for the proposed estimators and the test statistics. An extension to the multiple regressors with mixed persistency degrees is discussed in Section 3. Section 4 is for the Monte Carlo simulation studies and Section 5 reports the detailed results for an empirical application. Finally, Section 6 concludes the paper. The detailed proofs of the main asymptotic results are given in Appendix A.

Throughout this paper, the standard notations $\Rightarrow$, $\xrightarrow{d}$ and $\xrightarrow{p}$ are used to represent weak convergence and convergence in distribution as well as convergence in probability, respectively. All limits are for $T \to \infty$ in all theories, and $O_p(1)$ is stochastically asymptotically bounded while $o_p(1)$ is asymptotically negligible.

2. Econometric modeling procedures

2.1. Model framework

Assume that y_t is a dependent variable and its τth conditional quantile is $Q_{y_t}(\tau \mid \mathcal{F}_{t-1})$ such that $\Pr\left(y_t \le Q_{y_t}(\tau \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1}\right) = \tau \in (0,1)$, where $\mathcal{F}_{t-1}$ is the information set available at time $t-1$. For simplicity, a linear[2] predictive quantile regression is given by

$$Q_{y_t}(\tau \mid \mathcal{F}_{t-1}) = Q_{y_t}(\tau \mid x_{t-1}) = \mu_\tau + \beta_\tau x_{t-1}, \tag{1}$$

where x_{t-1} is a predictor to be the presentative (proxy) of $\mathcal{F}_{t-1}$, such as dividend–price ratio, earnings–price ratio, macroeconomic variable and so on, which is a time series, commonly modeled by an autoregressive (AR) model as

$$x_t = \rho x_{t-1} + v_t, \ \rho = 1 + c/T^\alpha, \ \ 1 \le t \le T, \tag{2}$$

where $\alpha = 0$ or 1 and $x_0 = o_p(\sqrt{T})$. Of course, a higher order AR model can be considered for x_t in (2). For simplicity, we begin with the univariate predictive quantile regression to illustrate the main idea in this paper. For x_t, the following typical types of persistency with different values of c and α are considered: (1) stationary (I0): $\alpha = 0$ and $|1+c| < 1$; (2) local to unit root (NI1): $\alpha = 1$ and $c < 0$; (3) unit root (I1): $c = 0$; (4) local to unity on the explosive side (LE): $\alpha = 1$ and $c > 0$.

Of course, it is interesting to consider the other cases as $0 < \alpha < 1$, corresponding to the so-called mildly integrated (MI) processes ($c < 0$) or mildly explosive (ME) processes ($c > 0$). The latter can be used to explore the mild economic or financial bubbles and other applications, see Phillips et al. (2015) and the references therein. As this paper mainly focuses on the methodology contribution, we focus on the current setting with $\alpha = 0$ or 1 and leave the MI and ME processes with $\alpha \in (0,1)$ as topics for future studies.[3]

Here, following Lee (2016), a general weakly dependent innovation structure of the linear process on $\{v_t\}$ in (2) is imposed and listed below.

A1. Assume that v_t follows a linear process given by $v_t = \sum_{j=0}^{\infty} F_{xj}\varepsilon_{t-j}$, where ε_t is a martingale difference sequence (MDS) with $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and $var(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \Sigma_\varepsilon$ for $\Sigma_\varepsilon > 0$ and $E\|\varepsilon_t\|^{2+\nu} < \infty$ for some $\nu > 0$. Here, $F_{x0} = I_K$, K is the dimension of x_t and $\sum_{j=0}^{\infty} j\|F_{xj}\| < \infty$ and $F_x(1) = \sum_{j=0}^{\infty} F_{xj} > 0$, where $F_x(z) = \sum_{j=0}^{\infty} F_{xj}z^j$. The variance matrix of v_t can be expressed as $\Omega_{vv} = \sum_{h=-\infty}^{\infty} E(v_t v_{t-h}^\top) = F_x(1)\Sigma_\varepsilon F_x(1)^\top$.

Remark 1. Assumption A.1 allows for linear process dependence for v_t and imposes a conditionally homoskedastic MDS condition for ε_t. Also, note that $K = 1$ for the univariate case.

[2] Of course, it would be interesting to investigate a nonlinear or nonparametric predictive quantile regression and it would be a future research topic.

[3] Note that Phillips and Lee (2016) extended the IVX approach to both MI and ME processes in the framework of predictive mean regression, and we conjecture that our method would still hold for MI and ME cases. However, the theoretical deviation to establish the limiting theory should be very challenging.

Define $u_{t\tau} \equiv y_t - Q_{y_t}(\tau|\mathcal{F}_{t-1})$, which is commonly called the quantile measurement error, similar to the measurement error in the predictive mean regression model, and define $\psi_\tau(u_{t\tau}) = \tau - 1(u_{t\tau} < 0)$. Under Assumption A.1, it is easy to verify that $P(u_{t\tau} \le 0|\mathcal{F}_{t-1}) = \tau$, $E(\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1}) = 0$, $E(\psi_\tau^2(u_{t\tau})|\mathcal{F}_{t-1}) = \tau(1-\tau)$ and the fourth moment $E[\psi_\tau(u_{t\tau})^4] = E\left\{E[\psi_\tau(u_{t\tau})^4|\mathcal{F}_{t-1}]\right\} = -3\tau^4 + 6\tau^3 - 4\tau^2 + \tau$. Further, define $\Sigma_{\psi_\tau v} = \sum_{h=-\infty}^{\infty} E[\psi_\tau(u_{t\tau})v_{t+h}] = F_x(1)E[\psi_\tau(u_{t\tau})\varepsilon_t]$. Under Assumption A.1, using the iterative law of expectation, one can show easily that $\Sigma_{\psi_\tau v} < \infty$. Then, similar to Lee (2016), the functional central limit theorem (FCLT) for $\{\psi_\tau(u_{t\tau}), v_t\}$ holds

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \begin{pmatrix} \psi_\tau(u_{t\tau}) \\ v_t \end{pmatrix} \Rightarrow \begin{pmatrix} B_{\psi_\tau}(r) \\ B_v(r) \end{pmatrix} = BM \begin{pmatrix} \tau(1-\tau) & \Sigma_{\psi_\tau v} \\ \Sigma_{\psi_\tau v} & \Omega_{vv} \end{pmatrix}, \tag{3}$$

where $[B_{\psi_\tau}(r), B_v(r)]^\top$ is a vector of Brownian motions. Furthermore, the local to unity limit law implies that $x_{\lfloor rT \rfloor}/\sqrt{T} \Rightarrow J_x^c(r)$ for $0 \le r \le 1$, where $J_x^c(r) = \int_0^r e^{(r-s)c} dB_v(s)$ with NI1, I1 and LE predictor; see Phillips (1987) and Lee (2016) for details.

Next, define $\lambda_{\tau,t} = \text{Corr}(\psi_\tau(u_{t\tau}), v_t)$ and assume that $\lambda_{\tau,t} = \lambda_\tau$ for simplicity. Then, similar to Campbell and Yogo (2006) for the predictive mean regression model, Lee (2016) seminally showed that the conventional t test statistic $t_{\hat{\beta}_\tau}$ of the predictive quantile regression with nonstationary predictor has the following asymptotic behavior

$$t_{\hat{\beta}_\tau} \Rightarrow \sqrt{1-\lambda_\tau^2}\, Z + \lambda_\tau \int \bar{J}_x^c(r) dB_x(r) \Big/ \sqrt{\Omega_{vv} \int \bar{J}_x^c(r)^2 dr},$$

where $\bar{J}_x^c(r) = J_x^c(r) - \int J_x^c(r) dr$ is the demeaned Ornstein–Uhlenbeck process and Z represents the standard normal distributions.[4] Clearly, λ_τ measures the degree for the so-called *embedded endogeneity* as in Campbell and Yogo (2006) for the predictive mean regression model. Therefore, the conventional test statistics in predictive quantile regression with the NI1, I1 and LE predictor x_t are invalid if $\lambda_\tau \neq 0$. Moreover, it is almost impossible to distinguish the difference between I0 and NI1, and/or between NI1 and I1, and so on, because it is extremely challenging to estimate consistently the nuisance parameter c and to test if the persistency α equals one or not. Thus, it is necessary to develop a unified inference method to avoid the mistake of making a false judgement about the persistency of predictors under a quantile framework.

Now, some regular assumptions on the conditional density of $u_{t\tau}$ are imposed, similar to those in Xiao (2009) and Lee (2016).

A2. (i) The sequence of conditional stationary probability density functions $\{f_{u_{t\tau},t-1}(\cdot)\}$ of $\{u_{t\tau}\}$ given $\mathcal{F}_{t-1}$ evaluated at zero satisfies a moment condition with a non-degenerate mean $f_{u_\tau}(0) = E(f_{u_{t\tau},t-1}(0)) > 0$ and $E(f_{u_{t\tau},t-1}^{\vartheta}(0)) < \infty$ for some $\vartheta > 1$.

(ii) For each t and $\tau \in (0, 1)$, $f'_{u_{t\tau},t-1}(x)$ is bounded with probability one around zero, i.e., $f'_{u_{t\tau},t-1}(\epsilon) < \infty$ and $f_{u_{t\tau},t-1}(\epsilon) < \infty$ almost surely for all $|\epsilon| < \eta$ for some $\eta > 0$.

Remark 2. As argued in Xiao (2009), the conditions in Assumption A.2 are quite standard and not restrictive. Under the above conditions, one can show that the Bahadur representation as in Theorem 1 (see later) holds and $f_{u_\tau}(0)$ can be estimated using observed data.

2.2. Estimation approach

Motivated by the VA approach of predictive mean regression studied by Elliott (2011) and Breitung and Demetrescu (2015), the following new approach is proposed for the predictive quantile regression. The model in (1) can be re-written as follows:

$$Q_{y_t}(\tau|x_{t-1}) = \mu_\tau + \beta_\tau x_{t-1} = \mu_\tau + \beta_\tau x_{t-1}^* + \gamma_\tau z_{t-1}, \tag{4}$$

where $x_{t-1}^* = x_{t-1} - z_{t-1}$ and z_{t-1} is an additional (auxiliary) variable which is chosen in Section 2.3 in detail. Note that $\gamma_\tau = \beta_\tau$ in (4) is used to construct weighted combined estimator for β_τ later. Clearly, μ_τ, β_τ and γ_τ in (4) can be estimated by running the following quantile regression

$$\hat{\theta}_\tau \equiv \left(\hat{\mu}_\tau, \hat{\beta}_\tau, \hat{\gamma}_\tau\right)^\top = \arg \min_{\mu_\tau, \beta_\tau, \gamma_\tau} \sum_{t=2}^{T} \rho_\tau\left(y_t - \mu_\tau - \beta_\tau x_{t-1}^* - \gamma_\tau z_{t-1}\right), \tag{5}$$

where $\rho_\tau(u) = u[\tau - 1(u < 0)]$ is the so-called check function in the literature. Note that the VA approach proposed by Breitung and Demetrescu (2015) uses only $\hat{\gamma}_\tau$, the estimator of the coefficient of the auxiliary variable z_t, to construct the test statistic in the predictive mean regression, and requires z_t to be an additional variable less persistent than x_t, in order to guarantee that the asymptotic distribution of the test statistic is irrelevant to the nuisance parameter c. If so, however, the corresponding test statistic suffers from the loss of power for the case with nonstationary x_t.

[4] $B_x(r)$ is defined by the orthogonal decomposition of Brownian motion as $dB_{\psi_\tau}(r) = dB_{\psi_\tau.x}(r) + \Sigma_{\psi_\tau v}\Omega_{vv}^{-1} dB_x(r)$, see Phillips (1989) for details.

To avoid this problem, the VA approach is improved in the following two aspects. First, a combined approach is used to construct the appropriate additional variable z_t, such that its persistency is always the same as that for the predictor x_t, while its key component is independent of x_t for NI1, I1 and LE cases. Second, a weighted combined estimator is proposed by using the coefficients of x^*_{t-1} and the additional variable z_t. With these two improvements, one can show that the test statistic based on the weighted estimator, after a self-normalization to eliminate the nuisance parameter c, can avoid not only the size distortion but also the loss of power with arbitrary persistency.

Next, it turns to the discussion on how to construct the weighted estimator for given z_t and then, elaborating the choice of z_t, which is presented in Section 2.3. As mentioned earlier, the identity $\gamma_\tau = \beta_\tau$ implies that it should be better to combine $\hat{\beta}_\tau$ and $\hat{\gamma}_\tau$ together to obtain a weighted estimation for β_τ. Consequently, the rotation idea in the principle component analysis is applied to constructing the estimator for β_τ, which is the weighted sum of $\hat{\beta}_\tau$ and $\hat{\gamma}_\tau$, denoted by $\hat{\beta}^w_\tau$,

$$\hat{\beta}^w_\tau = \frac{W_1}{W_1 + W_2}\hat{\beta}_\tau + \frac{W_2}{W_1 + W_2}\hat{\gamma}_\tau, \tag{6}$$

where W_1 and W_2 are two weighting functions. By selecting some appropriate weights W_1 and W_2, one can construct a $\hat{\beta}^w_\tau$, whose asymptotic distribution follows a mixture normal distribution[5] and is irrelevant to the nuisance parameter c after normalization. For this purpose, the weights W_1 and W_2 are taken to be

$$W_1 = \sum_{t=2}^{T} x^*_{t-1} z_{t-1}/T^2 - \sum_{t=2}^{T} x^*_{t-1} \sum_{t=2}^{T} z_{t-1}/T^3, \tag{7}$$

and

$$W_2 = \sum_{t=2}^{T} z^2_{t-1}/T^2 - \left(\sum_{t=2}^{T} z_{t-1}\right)^2 /T^3. \tag{8}$$

In Section 2.4, more detailed arguments will be provided to explain the reason on why the above W_1 and W_2 are used.

2.3. Choice of auxiliary variable

This subsection is devoted to constructing the additional regressor z_{t-1}, such that our method is valid for both stationary and nonstationary predictor without sacrificing any convergence rate. To achieve this target, a three-step approach is proposed to construct z_{t-1}. First, an exogenous unit root process $\zeta_{t-1} = \sum_{s=1}^{t-1} \varsigma_s$ is generated, where $\varsigma_s \sim iid(0, 1)$. Therefore, $W_{\zeta,T}(\cdot) \Rightarrow B(\cdot)$ based on the FCLT, where $W_{\zeta,T}(r) = \zeta_{\lfloor rT \rfloor}/\sqrt{T}$ for $0 \le r \le 1$ and $B(\cdot)$ is the standard Brownian motion. Second, obtain the ordinary least squared estimator $\hat{\pi}_1$ from the following regression

$$x_{t-1} = \pi_0 + \pi_1 \zeta_{t-1} + e_t. \tag{9}$$

Third, define z_{t-1} as a linear combination of ζ_{t-1} and a bounded transformation of x_{t-1} as

$$z_{t-1} = \hat{\pi}_1 \zeta_{t-1} + x_{t-1}/\sqrt{1 + x^2_{t-1}}. \tag{10}$$

Note that the second term in the above equation $x_{t-1}/\sqrt{1 + x^2_{t-1}}$ is always bounded with probability 1 for any stationary and nonstationary x_{t-1}.

Remark 3. Indeed, the idea of using an independent random walk process as the instrumental variable is similar to that in Breitung and Demetrescu (2015) under the framework of predictive mean regressions, by considering two types of instruments: Type-I and Type-II instruments. Type-I instruments are generated from the original predictor x_{t-1} but are required to be less persistent than x_{t-1}. A special case of Type-I instruments is the mild integrated instrument variable adopted in the IVX approach in Kostakis et al. (2015). Type-II instruments include strictly exogenous nonstationary variables, deterministic terms and Cauchy type instrument. Therefore, in this sense, both ζ_{t-1} and $x_{t-1}/\sqrt{1 + x^2_{t-1}}$ can be regraded as Type-II instruments, as $x_{t-1}/\sqrt{1 + x^2_{t-1}}$ converges to the Cauchy instrument $sign(x_{t-1})$ for nonstationary x_{t-1}, see Choi et al. (2016). However, the random walk instrument ζ_{t-1} does not work for stationary cases, while $x_{t-1}/\sqrt{1 + x^2_{t-1}}$ cannot handle the predictive regression with intercept term for nonstationary cases without some necessary adjustments.[6]

5 For the definition of mixture normal, the reader is referred to the paper by Phillips (1987). That is, $Y \sim MN(\mu, \Sigma)$ means $Y \sim N(\mu, \Sigma)$ given μ and Σ, which might be random.

6 In predictive mean regressions with intercept term, Zhu et al. (2014) and Liu et al. (2019) applied the sample splitting approach to remove the impact of intercept, with a loss of information. However, the sample splitting approach does not work in the quantile regression framework.

Here, we take a weighted combination of ζ_{t-1} and $x_{t-1}/\sqrt{1+x_{t-1}^2}$, with the weight $\hat{\pi}_1$ estimated from (9). By doing so, our method is robust to both nonstationary and stationary cases, and can be easily extended to the multivariate case with mixed persistency.

Remark 4. Note that the form of bounded transformation of x_{t-1} might not be unique. One could define z_{t-1} as a linear combination of ζ_{t-1} and a general bounded transformation $g(x_{t-1})$ like $z_{t-1} = \hat{\pi}_1\zeta_{t-1}+g(x_{t-1})$. To guarantee the validity of our approach, $g(x_{t-1})$ just needs to satisfy two conditions. First, for nonstationary x_{t-1}, the transformation $g(x_{t-1})$ should be bounded. Second, for stationary x_{t-1}, $Var\left[g(x_{t-1})\right]$ exists and deviates from zero, i.e., $g(x_{t-1})$ is not degenerate, so the central limit theory can be applied.[7]

From the regression (9), it is easy to establish the asymptotic properties of $\hat{\pi}_1$ for the cases with stationary and nonstationary x_t, respectively. For nonstationary x_{t-1}, it is easy to show that

$$\hat{\pi}_1 = \sum_{t=2}^{T} \bar{x}_{t-1}\bar{\zeta}_{t-1}/\sum_{t=2}^{T}\bar{\zeta}_{t-1}^2 = \int \bar{B}(r)\bar{J}_x^c(r)dr/\int \bar{B}(r)^2 dr + o_p(1) = \tilde{\pi}_1 + o_p(1),$$

where $\bar{x}_{t-1} = x_{t-1} - \sum_{t=2}^{T} x_{t-1}/T$, $\bar{\zeta}_{t-1} = \zeta_{t-1} - \sum_{t=2}^{T}\zeta_{t-1}/T$, and $\tilde{\pi}_1$ is a nonzero random variable due to the spurious correlation between x_{t-1} and ζ_{t-1} similar to that in Phillips (2014). For stationary x_{t-1}, $\hat{\pi}_1$ converges to zero with the convergence rate T, i.e., $\hat{\pi}_1 = O_p(1/T)$.

Given the above asymptotic results of $\hat{\pi}_1$, one can show that, for nonstationary cases, the second term $x_{t-1}/\sqrt{1+x_{t-1}^2}$ in Eq. (10) is dominated by the first term $\hat{\pi}_1\zeta_{t-1}$, while for stationary cases, the first term is dominated by the second term $x_{t-1}/\sqrt{1+x_{t-1}^2}$. In certain sense, the coefficient $\hat{\pi}_1$ plays a role of filtering such that the auxiliary variable z_{t-1} has the same persistency as x_{t-1} does, and that is why our method can achieve the optimal convergence rate for both stationary and nonstationary cases.

Next, we can establish the asymptotic property of $W_1 + W_2$, which is useful to derive the main results of the paper. For nonstationary x_{t-1}, as z_{t-1} is determined by $\hat{\pi}_1\zeta_{t-1}$, can show easily that

$$W_1 + W_2 = \hat{\pi}_1 \sum_{t=2}^{T} \bar{x}_{t-1}\bar{\zeta}_{t-1}/T^2 + o_p(1) = \tilde{\pi}_1 \int \bar{B}(r)\bar{J}_x^c(r)dr + o_p(1). \tag{11}$$

For stationary x_{t-1}, z_{t-1} is determined by $x_{t-1}/\sqrt{1+x_{t-1}^2}$, and then,

$$T(W_1+W_2) = \frac{1}{T}\sum_{t=2}^{T}\bar{x}_{t-1}x_{t-1}/\sqrt{1+x_{t-1}^2} + o_p(1) = E\left[x_t^2(1+x_t^2)^{-1/2}\right] + o_p(1). \tag{12}$$

2.4. Large sample theory

To obtain the asymptotic distribution of $\hat{\beta}_\tau^w$, first, we establish the so-called Bahadur representation for $\hat{\theta}_\tau$, which is commonly used for stationary quantile regression to get an explicit expression of estimators; see, for example, Cai and Xu (2008) for details. To this end, denote $\theta_\tau = (\mu_\tau, \beta_\tau, \gamma_\tau)^\top$ as the vector of true values of coefficients in Eq. (4). Also, define the weighting matrix $\mathbf{D}_T = \text{diag}(\sqrt{T}, T, T)$ for NI1, I1 and LE, and $\mathbf{D}_T = \text{diag}(\sqrt{T}, \sqrt{T}, \sqrt{T})$ for I0. The following theorem states the Bahadur representation for $\mathbf{D}_T(\hat{\theta}_\tau - \theta_\tau)$, and its mathematical proof is given in Appendix A. Note that this result is new in the literature when regressors might be nonstationary and is of own interest.

Theorem 1 (*Bahadur Representation*). *Under Assumptions A.1 and A.2,*

$$\mathbf{D}_T(\hat{\theta}_\tau - \theta_\tau) = f_{u_\tau}(0)^{-1} N_T^{-1}\mathbf{D}_T^{-1}\sum_{t=2}^{T}\Lambda_{t-1}\psi_\tau(u_{t\tau}) + o_p(1), \tag{13}$$

where $\Lambda_{t-1} = (1, x_{t-1}^*, z_{t-1})^\top$, $N_T = \mathbf{D}_T^{-1}\sum_{t=2}^{T}\Lambda_{t-1}\Lambda_{t-1}^\top\mathbf{D}_T^{-1}$, *and* $f_{u_\tau}(0)$ *is defined in Assumption A.2.*

Remark 5. From Theorem 1, one can see clearly that the right-hand side of (13) still involves x_{t-1}^*, leading to a nonstandard distortion in the asymptotic distribution if $\lambda_\tau \neq 0$, see Lee (2016). To construct a pivotal test statistic free of nuisance

[7] To check whether our method is sensitive to the forms of the transformation, we conduct simulations with other bounded transformations, and find that the performance of our test is quite similar. To save space, we skip to report the simulation results for this study in the paper, but the codes and results are available upon request.

parameter c, the weighted estimator $\hat{\beta}_\tau^w$ is constructed as in (6), with a similar idea of rotation in factor analysis, to get rid of the impact of x_{t-1}^*. It will then be shown by (A.8) in Appendix A that the following result holds for $\hat{\beta}_\tau^w$,

$$(W_1 + W_2)T(\hat{\beta}_\tau^w - \beta_\tau) = f_{u_\tau}(0)^{-1} \sum_{t=2}^{T} \frac{1}{\sqrt{T}} \left(z_{t-1} - \sum_{t=2}^{T} z_{t-1}/T \right) \psi_\tau(u_{t\tau})/\sqrt{T} + o_p(1). \tag{14}$$

Evidently, the right-hand side of (14) involves only z_{t-1} but not x_{t-1} or x_{t-1}^*, so that it makes the asymptotic (or mixture) normality of $\hat{\beta}_\tau^w$ dependent only on z_{t-1}.

Remark 6. It is very interesting to see whether W_1 and W_2 could be defined under some other criteria, for instance, maximizing the local power. As this paper focuses on the robust inference across different types of persistent predictors, our target is to obtain a test statistic converging to a standard limiting distribution without any nuisance parameters. We therefore choose W_1 and W_2 as (7) and (8) to remove the impact of the nonstandard distortion in the limiting distribution of the weighted estimator $\hat{\beta}_\tau^w$.

Next, one of the main results in this paper is stated in the following theorem with its detailed proof given in Appendix A.

Theorem 2. *Under Assumptions A.1 and A.2, for I0, NI1, I1 and LE cases, the asymptotic distribution of $\hat{\beta}_\tau^w$ is given as follows*

$$\begin{cases} \sqrt{T}(\hat{\beta}_\tau^w - \beta_\tau) \xrightarrow{d} \mathrm{N}\left(0, \sigma_{\beta_\tau}^2\right), & I0, \\ T\,\pi_c\,(\hat{\beta}_\tau^w - \beta_\tau) \xrightarrow{d} \mathrm{N}\left(0, \sigma_\tau^2\right), & NI1,\ I1 \text{ and } LE, \end{cases}$$

where $\sigma_\tau^2 = \tau(1-\tau)/f_{u_\tau}^2(0)$, $\sigma_{\beta_\tau}^2 = \sigma_\tau^2 \left\{E\left[x_t^2(1+x_t^2)^{-1/2}\right]\right\}^{-2} Var\left[x_t(1+x_t^2)^{-1/2}\right]$ and $\pi_c = \int \bar{B}(r)\bar{J}_x^c(r)dr \left[\int \bar{B}^2(r)dr\right]^{-1/2}$.

Remark 7. Clearly, Theorem 2 shows the convergence rate of the estimator of $\hat{\beta}_\tau^w$ with N1, I1 and LE x_t is faster than that for the IVX-QR method proposed in Lee (2016).

By a self normalization, we can construct the following t-test statistic t^w:

$$t^w = \hat{f}_{u_\tau}(0)\,[W_2\tau(1-\tau)]^{-1/2}\,(W_1 + W_2)T\hat{\beta}_\tau^w,$$

where $\hat{f}_{u_\tau}(0)$ is a consistent estimator of $f_{u_\tau}(0)$ and the detailed construction of $\hat{f}_{u_\tau}(0)$ can be found in Lee (2016). The following theorem states the asymptotic behavior of the proposed t-test statistic t^w under both the null hypothesis and the local alternative hypothesis with its detailed proof delegated to Appendix A.

Theorem 3. *(1) Under the null hypothesis $H_0 : \beta_\tau = 0$, t^w converges to the standard normal. (2) Under the local alternative hypothesis $H_a : \beta_\tau = b_\tau/T^{(1+\alpha)/2}$ for any b_τ, t^w converges to the standard normal plus a constant $b_\tau/\sigma_{\beta_\tau}$, if x_{t-1} is I0, and it converges to the standard normal plus a random variable $b_\tau|\pi_c|/\sigma_\tau$, if x_{t-1} is NI1, I1 or LE, where σ_{β_τ}, π_c and σ_τ are defined in Theorem 2.*

Remark 8. From Theorem 3, one can conclude that the test statistic t^w reaches the optimal convergence rate T for NI1, I1 and LE predictor x_{t-1} and $\sqrt{T}$ for I0 predictor x_{t-1}. In particular, for nonstationary case, the quantity $b_\tau|\pi_c|/\sigma_\tau$, the deviation from the standard normality, varies between $(-\infty, 0)$ or $(0, +\infty)$, relying on the sign of b_τ only. Thus, t^w enjoys an additional increase of local power compared to the t-test statistic in Breitung and Demetrescu (2015), where its local lower relies on a deviation varying between $(-\infty, +\infty)$, see Part 1 of Corollary 3 and Remark 4 in Breitung and Demetrescu (2015).

3. Multiple predictive quantile regressions

When some of regressors are nonstationary and some are stationary in a multiple regression, it is well known in the literature that the convergence rates for estimators of coefficients are totally different for nonstationary and stationary regressors; see, for example, Cai and Wang (2014). When regressors are nonstationary, as pointed out by Phillips and Lee (2013), the Bonferroni's method in Campbell and Yogo (2006) and the weighted empirical likelihood approach in Zhu et al. (2014), Liu et al. (2019), and Yang et al. (2019) cannot be easily extended to multiple regressions. In contrast, our method can be easily extended to multivariate predictive quantile regressions with mixed persistency.

Particularly, we consider the following multivariate predictive quantile regression model:

$$Q_{y_t}(\tau|\mathbf{X}_{t-1}) = \mu_\tau + \boldsymbol{\beta}_\tau^\top \mathbf{X}_{t-1}, \tag{15}$$

where $\boldsymbol{\beta}_\tau = (\beta_{1\tau}, \beta_{2\tau}, \ldots, \beta_{K\tau})^\top$ is a $K \times 1$ vector and $\mathbf{X}_{t-1}$ is a $K \times 1$ vector of predictors, which might contain both stationary and nonstationary predictors. To this end, $\mathbf{X}_{t-1}$ is written as $\mathbf{X}_{t-1} = (\mathbf{X}_{1,t-1}^\top, \mathbf{X}_{2,t-1}^\top)^\top$ with $\mathbf{X}_{1,t-1} = (x_{1,t-1}, x_{2,t-1} \cdots, x_{K_1,t-1})^\top$ being nonstationary and $\mathbf{X}_{2,t-1} = (x_{K_1+1,t-1}, x_{K_1+2,t-1}, \ldots, x_{K,t-1})^\top$ being stationary. It is assumed there is no cointegration relationship among $\mathbf{X}_{1,t-1}$. Note that $0 \le K_1 \le K$ and $K_1 = 0$ means all elements in $\mathbf{X}_{t-1}$ are I0, while $K_1 = K$ means all elements in $\mathbf{X}_{t-1}$ are NI1, I1 or LE. Now, $x_{i,t}$ can be modeled by an AR(1) as $x_{i,t} = \rho_i x_{i,t-1} + v_{i,t}$, where $\rho_i = 1 + c_i/T$ for $1 \le i \le K_1$ and $\rho_i = 1 + c_i$ with $|1 + c_i| < 1$ for $K_1 + 1 \le i \le K$ for all $1 \le t \le T$. Thus, different predictors in multivariate predictive quantile regression are allowed to have different degrees of persistency. Similar to the univariate case, the local to unity limit law holds for all nonstationary predictors and for $i = 1, \ldots, K_1$, $x_{i,\lfloor rT\rfloor}/\sqrt{T} \Rightarrow J_{x_i}^{c_i}(r)$ and $J_{x_i}^{c_i}(r) = \int_0^r e^{(r-s)c_i} dB_{v_i}(s)$, where $B_{v_i}(s)$ is the ith element of $B_v(s)$, which is a vector of Brownian motions defined in (3).

Remark 9. The model in (15) covers some known models in mean models in the literature. For example, if there is nonstationary part ($K_1 = 0$), (15) reduces to the model studied by Amihud et al. (2009) for mean regression models.

To estimate μ_τ and $\boldsymbol{\beta}_\tau$ in (15), let $\mathbf{X}_{t-1}^* = \mathbf{X}_{t-1} - \mathbf{Z}_{t-1}$ and $\mathbf{Z}_{t-1}$ be the vector of additional variables. Then, μ_τ and $\boldsymbol{\beta}_\tau$ can be estimated based on the VA as follows:

$$\left(\hat{\mu}_\tau, \hat{\boldsymbol{\beta}}_\tau, \hat{\gamma}_\tau\right)^\top = \arg\min_{\mu_\tau, \boldsymbol{\beta}_\tau, \gamma_\tau} \sum_{t=2}^{T} \rho_\tau\left(y_t - \mu_\tau - \boldsymbol{\beta}_\tau^\top \mathbf{X}_{t-1}^* - \gamma_\tau^\top \mathbf{Z}_{t-1}\right),$$

where $\mathbf{Z}_t = (z_{1,t}, z_{2,t}\cdots, z_{K,t})^\top$ is constructed by three steps similar to the univariate case as in Section 2.3; that is, first, for each i, $\zeta_{i,t-1} = \sum_{s=1}^{t-1} \varsigma_{i,s}$, where $\varsigma_{i,s} \sim iid(0,1)$ generated by simulation and thus, independent of y_t and $\mathbf{X}_t$. Therefore, $W_{i,\zeta,T}(\cdot) \Rightarrow B_i(\cdot)$ based on the FCLT, where $W_{i,\zeta,T}(r) = \zeta_{i,rT}/\sqrt{T}$ for $0 \le r \le 1$ and $B_i(\cdot)$ is the standard Brownian motion. Secondly, for each $1 \le i \le K$, run the regression $x_{i,t} = \pi_{0,i} + \pi_{1,i}\zeta_{i,t-1} + e_{i,t}$ to obtain the ordinary least squared estimator $\hat{\pi}_{1,i}$. Similarly, one can show that $\hat{\pi}_{1,i} \xrightarrow{d} \tilde{\pi}_{1,i} = \int \bar{B}_i(r) J_{x_i}^{c_i}(r)dr / \int \bar{B}_i(r)^2 dr$, where $\bar{B}_i(r) = B_i(r) - \int B_i(r)dr$ for nonstationary $x_{i,t}$ while $\hat{\pi}_{1,i} = O_p(T^{-1})$ for stationary $x_{i,t}$. Thirdly, we define $z_{i,t-1}$ as a linear combination of $\zeta_{i,t-1}$ and one bounded transformation of $x_{i,t-1}$ as $z_{i,t-1} = \hat{\pi}_{1,i}\zeta_{i,t-1} + x_{i,t-1}/\sqrt{1 + x_{i,t-1}^2}$. Since the procedure could be implemented one predictor by one predictor and each step does not rely on others, then our proposed method is valid in multivariate predictive quantile regression with mixed persistency.

Similar to the univariate case, the weighted estimator $\hat{\boldsymbol{\beta}}_\tau^w$ in the multivariate predictive quantile regression is given as follows:

$$\hat{\boldsymbol{\beta}}_\tau^w = (\mathbf{W}_1 + \mathbf{W}_2)^{-1}\left(\mathbf{W}_1 \hat{\boldsymbol{\beta}}_\tau + \mathbf{W}_2 \hat{\gamma}_\tau\right),$$

where $\mathbf{W}_1 = \sum_{t=2}^T \mathbf{Z}_{t-1}(\mathbf{X}_{t-1}^*)^\top/T^2 - \sum_{t=2}^T \mathbf{Z}_{t-1}\sum_{t=2}^T(\mathbf{X}_{t-1}^*)^\top/T^3$ and $\mathbf{W}_2 = \sum_{t=2}^T \mathbf{Z}_{t-1}\mathbf{Z}_{t-1}^\top/T^2 - \sum_{t=2}^T \mathbf{Z}_{t-1}\sum_{t=2}^T \mathbf{Z}_{t-1}^\top/T^3$. Without loss of generalization, the asymptotic property of $\hat{\boldsymbol{\beta}}_\tau^w$ is presented for the special case with $K = 2$ in the following theorem. For different mixed persistency cases, we define the weighting matrix $\tilde{D}_T$ as follows: $\tilde{D}_T = \mathrm{diag}(\sqrt{T}, \sqrt{T})$, if $K_1 = 0$; $\tilde{D}_T = \mathrm{diag}(T, \sqrt{T})$, if $K_1 = 1$; $\tilde{D}_T = \mathrm{diag}(T, T)$, if $K_1 = 2$.

Furthermore, to describe the asymptotic properties for $\hat{\boldsymbol{\beta}}_\tau^w$, we define the following two matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ for three cases as follows:

Case 1 ($K_1 = 0$):

$$\mathbf{V}_1 = \begin{pmatrix} E\left(x_{1,t}^2/\sqrt{1+x_{1,t}^2}\right) & E\left(x_{1,t}x_{2,t}/\sqrt{1+x_{1,t}^2}\right) \\ E\left(x_{1,t}x_{2,t}/\sqrt{1+x_{2,t}^2}\right) & E\left(x_{2,t}^2/\sqrt{1+x_{2,t}^2}\right) \end{pmatrix}, \tag{16}$$

and

$$\mathbf{V}_2 = \begin{pmatrix} \mathrm{Var}\left(x_{1,t}/\sqrt{1+x_{1,t}^2}\right) & \mathrm{Cov}\left(x_{1,t}/\sqrt{1+x_{1,t}^2}, x_{2,t}/\sqrt{1+x_{2t}^2}\right) \\ \mathrm{Cov}\left(x_{1,t}/\sqrt{1+x_{1,t}^2}, x_{2,t}/\sqrt{1+x_{2t}^2}\right) & \mathrm{Var}\left(x_{2,t}/\sqrt{1+x_{2,t}^2}\right) \end{pmatrix}. \tag{17}$$

Case 2 ($K_1 = 1$):

$$\mathbf{V}_1 = \mathrm{diag}\left\{\tilde{\pi}_{1,1}\int \bar{B}_1(r)J_{x_1}^{c_1}(r)dr,\ \ E\left(x_{2,t}^2/\sqrt{1+x_{2,t}^2}\right)\right\}, \tag{18}$$

and

$$\mathbf{V}_2 = \mathrm{diag}\left\{\tilde{\pi}_{1,1}^2\int \bar{B}_1(r)^2\,dr,\ \ \mathrm{Var}\left(x_{2,t}/\sqrt{1+x_{2,t}^2}\right)\right\}. \tag{19}$$

Case 3 ($K_1 = 2$):

$$\mathbf{V}_1 = \begin{pmatrix} \tilde{\pi}_{1,1}\int \bar{B}_1(r)J_{x_1}^{c_1}(r)dr & \tilde{\pi}_{1,1}\int \bar{B}_1(r)J_{x_2}^{c_2}(r)dr \\ \tilde{\pi}_{1,2}\int \bar{B}_2(r)J_{x_1}^{c_1}(r)dr & \tilde{\pi}_{1,2}\int \bar{B}_2(r)J_{x_2}^{c_2}(r)dr \end{pmatrix}, \tag{20}$$

and

$$\mathbf{V}_2 = \begin{pmatrix} \tilde{\pi}_{1,1}^2 \int \bar{B}_1(r)^2 dr & \tilde{\pi}_{1,1}\tilde{\pi}_{1,2} \int \bar{B}_1(r)\bar{B}_2(r)dr \\ \tilde{\pi}_{1,1}\tilde{\pi}_{1,2} \int \bar{B}_2(r)\bar{B}_1(r)dr & \tilde{\pi}_{1,2}^2 \int \bar{B}_2(r)^2 dr \end{pmatrix}. \tag{21}$$

Then, the asymptotic distribution for $\hat{\boldsymbol{\beta}}_\tau^w$ is stated in the following theorem with its proof delegated to Appendix A.

Theorem 4. *Under Assumptions A.1 and A.2, the asymptotic distribution of $\hat{\boldsymbol{\beta}}_\tau^w$ is given by*

$$\tilde{D}_T(\hat{\boldsymbol{\beta}}_\tau^w - \boldsymbol{\beta}_\tau) \xrightarrow{d} f_{u_\tau}(0)^{-1}\mathbf{V}_1^{-1}\,\mathrm{MN}\,(0, \tau(1-\tau)\mathbf{V}_2)\,,$$

where $\bar{\mathbf{Z}}_{t-1} = \mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}$, $\mathbf{V}_1$ and $\mathbf{V}_2$ are defined in (16)–(21), respectively.

To test $H_0 : \mathbf{R}\boldsymbol{\beta}_\tau = \mathbf{r}_\tau$, where $\mathbf{R}$ is a $r \times K$ matrix with the rank r, a Wald type test statistic Q_m^w can be easily constructed as follows:

$$Q_m^w = \frac{\hat{f}_{u_\tau}(0)^2}{\tau(1-\tau)}T^2(\mathbf{R}\hat{\boldsymbol{\beta}}_\tau^w - \mathbf{r}_\tau)^\top \left\{\mathbf{R}(\mathbf{W}_1+\mathbf{W}_2)^{-1}\mathbf{W}_2\left[\mathbf{R}(\mathbf{W}_1+\mathbf{W}_2)^{-1}\right]^\top\right\}^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}_\tau^w - \mathbf{r}_\tau),$$

where $\hat{f}_{u_\tau}(0)$ is a consistent estimator of $f_{u_\tau}(0)$. The limiting distribution of Q_m^w under the null hypothesis is stated in the following theorem with its detailed proof given in Appendix A.

Theorem 5. *Under Assumptions A.1 and A.2 and the null hypothesis H_0: $R\boldsymbol{\beta}_\tau = \mathbf{r}_\tau$, one can show that the limiting distribution of Q_m^w is a χ^2-distribution with r degrees of freedom.*

Remark 10. Theorem 5 implies that our method can be applied to test a general linear hypothesis of $\boldsymbol{\beta}_\tau$, for example, to test the predicability of one specific predicting variable or a subset of predicting variables.

4. Monte Carlo simulations

To demonstrate the effectiveness of the proposed method, two Monte Carlo simulation experiments are considered. The first experiment considers a data generating process (DGP) with a univariate predictor, while the second experiment is devoted to a bivariate case with mixed persistences. For each experiment, we conduct a comparison between our method and the IVX-QR approach, except for the single test in the bivariate case.

Example 1. In this example, the following DGP is set up for the univariate quantile regression:

$$y_t = 3(\mu + \beta x_{t-1}) + (\mu + \beta x_{t-1})u_t, \quad \text{and} \quad x_t = \rho x_{t-1} + v_t,$$

where $\mu = 1$ and $\rho = 1 + c/T^\alpha$. To create the embedded endogeneity among innovations, the innovation processes are generated as $(u_t, v_t)^\top \sim iid\,N(0_{2\times1}, \Sigma_{2\times2})$, where $\Sigma = \begin{pmatrix} 1 & -0.95 \\ -0.95 & 1 \end{pmatrix}$. By Proposition 1 of Gaglianone et al. (2011), it is easy to see that the conditional quantile of y_t given x_{t-1} at the quantile level τ is given by

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = Q_{u_t}(\tau) + 3 + \beta[Q_{u_t}(\tau) + 3]x_{t-1} = \mu_\tau + \beta_\tau x_{t-1},$$

where $\mu_\tau = Q_{u_t}(\tau) + 3$, $\beta_\tau = \beta[Q_{u_t}(\tau) + 3]$ and $Q_{u_t}(\tau)$ is the τth quantile of u_t.[8]

We first compare the size performance of our method and the IVX-QR approach. Similar to Lee (2016), the IVX-QR based on the simple QR testing procedure is implemented with the filtering parameters chosen according to the practical rule suggested by Lee (2016). The nominal size is set at 5% and the sample size T is set at 700. We consider a large sample size as long time series data are often available in financial applications. Note that we also conduct simulations with sample sizes $T = 150$ and 300 and conclusions are similar so that they are not presented here to save space and available upon request. For nonstationary setting, the values of c are chosen from $(1.5, 0, -5, -25)$, corresponding to the LE, I1, NI1 and NI1 (with large deviation from unit root) processes. For stationary setting, the values of c are chosen from $(-0.05, -0.1, -0.15, -0.2)$, corresponding to the stationary AR(1) processes with $\rho = (0.95, 0.9, 0.85, 0.8)$. For each setting, the rejection rate is calculated based on 500 simulations, and we repeat 100 times of simulations and calculations. The mean and the standard error in parenthesis of 100 rejection rates are then reported in Panel A and B of Table 1 for the nonstationary case ($\alpha = 1$) and the stationary case ($\alpha = 0$), given different values of the persistency parameter c and at different quantile levels τ.

The following findings can be evidently observed from Table 1. First, our method shows better size performance compared to the IVX-QR approach for most cases. Second, for quantile level τ close to 0.5, the size of the proposed method is very close to the nominal size at 0.05, while IVX-QR tends to over-reject for the case of LE predictors. Third, due to less data points in tails, both methods have size distortions for the extreme quantile levels.

[8] Following Maynard et al. (2011), we set the DGP process as a random coefficient model to allow the impact of x_{t-1} to vary across the quantiles of u_t. We also conduct simulations using the same DGP of Lee (2016), which is a location shift model with a fixed β_τ, and our main conclusions are similar.

Table 1
Size performances with the nominal size 5% and T = 700.

Panel A: $\alpha = 1$												
	τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
t^w	$c = 1.5$	0.069	0.059	0.051	0.049	0.048	0.047	0.047	0.049	0.051	0.060	0.071
		(0.012)	(0.010)	(0.009)	(0.009)	(0.010)	(0.011)	(0.009)	(0.009)	(0.011)	(0.010)	(0.011)
	$c = 0$	0.069	0.061	0.054	0.050	0.050	0.051	0.051	0.052	0.054	0.061	0.070
		(0.011)	(0.011)	(0.012)	(0.010)	(0.009)	(0.010)	(0.012)	(0.009)	(0.010)	(0.011)	(0.012)
	$c = -5$	0.068	0.058	0.052	0.049	0.047	0.046	0.046	0.048	0.051	0.060	0.069
		(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.012)
	$c = -15$	0.070	0.059	0.053	0.051	0.050	0.049	0.049	0.050	0.052	0.060	0.071
		(0.011)	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)	(0.011)
IVX-QR	$c = 1.5$	0.147	0.142	0.117	0.146	0.104	0.103	0.104	0.143	0.115	0.140	0.146
		(0.015)	(0.015)	(0.015)	(0.015)	(0.014)	(0.013)	(0.014)	(0.014)	(0.015)	(0.015)	(0.014)
	$c = 0$	0.103	0.094	0.070	0.093	0.059	0.059	0.059	0.092	0.070	0.094	0.102
		(0.012)	(0.012)	(0.011)	(0.013)	(0.009)	(0.010)	(0.011)	(0.012)	(0.011)	(0.012)	(0.013)
	$c = -5$	0.071	0.059	0.046	0.050	0.041	0.041	0.042	0.049	0.044	0.057	0.069
		(0.011)	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.012)
	$c = -25$	0.088	0.073	0.059	0.056	0.055	0.054	0.054	0.057	0.060	0.073	0.089
		(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.010)	(0.008)	(0.010)	(0.011)	(0.012)	(0.012)
Panel B: $\alpha = 0$												
t^w	$c = -0.05$	0.073	0.064	0.058	0.055	0.050	0.052	0.051	0.054	0.055	0.063	0.074
		(0.013)	(0.010)	(0.012)	(0.010)	(0.010)	(0.011)	(0.010)	(0.009)	(0.010)	(0.012)	(0.011)
	$c = -0.1$	0.075	0.064	0.056	0.055	0.052	0.051	0.052	0.054	0.056	0.064	0.073
		(0.011)	(0.011)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.013)
	$c = -0.15$	0.074	0.063	0.055	0.053	0.052	0.051	0.053	0.052	0.053	0.062	0.073
		(0.013)	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.008)	(0.009)	(0.012)	(0.011)
	$c = -0.2$	0.075	0.063	0.055	0.053	0.051	0.052	0.052	0.051	0.054	0.060	0.074
		(0.013)	(0.010)	(0.009)	(0.010)	(0.011)	(0.010)	(0.009)	(0.009)	(0.011)	(0.011)	(0.012)
IVX-QR	$c = -0.05$	0.067	0.054	0.045	0.044	0.040	0.040	0.040	0.043	0.044	0.051	0.066
		(0.010)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)
	$c = -0.1$	0.067	0.053	0.046	0.044	0.042	0.042	0.040	0.043	0.045	0.053	0.067
		(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.008)	(0.010)	(0.009)	(0.010)	(0.010)
	$c = -0.15$	0.069	0.055	0.047	0.045	0.042	0.042	0.042	0.043	0.045	0.054	0.068
		(0.011)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)
	$c = -0.2$	0.070	0.056	0.047	0.045	0.042	0.042	0.042	0.044	0.046	0.054	0.070
		(0.013)	(0.011)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.008)	(0.009)	(0.011)

Note: The DGP is given by $y_t = 3(1 + \beta x_{t-1}) + (1 + \beta x_{t-1})u_t$ where $\beta = 0$ and $x_t = \rho x_{t-1} + v_t$ with $\rho = 1 + c/T^{\alpha}$. The innovation processes are generated as $(u_t, v_t)^\top \sim iid\, N(0_{2\times 1}, \Sigma_{2\times 2})$, where $\Sigma = \begin{pmatrix} 1 & -0.95 \\ -0.95 & 1 \end{pmatrix}$.

Next, we conduct a comparison of the power of the proposed method with that for the IVX-QR method. To this end, at the nominal size 5%, Figs. 1 and 2 display the results for $\alpha = 1$ and $\alpha = 0$ with different c, given $\tau = 0.5$ and the sample size $T = 700$. To see the local power, we set $\beta = b/T^{(1+\alpha)/2}$ and thus, $\beta_\tau = b_\tau/T^{(1+\alpha)/2} = b[Q_{u_t}(\tau) + 3]/T^{(1+\alpha)/2}$. Evidently, our method performs better than the IVX-QR method in terms of power for all cases. This finding confirms the theory in Theorem 3, which states that the convergence rate of the newly proposed method is faster than that for the IVX-QR method. Note that we also conduct the simulations for sample size $T = 300$ and $\tau = 0.05$, 0.5 and 0.95 and obtain similar conclusions, omitted here to save space and available upon request.

Example 2. In this example, we consider a bivariate quantile regression model with mixed persistences. The DGP is set up as follows:

$$y_t = 3(\mu + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1}) + (\mu + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1})u_t,$$

where $\mu = 1$, $x_{1,t} = (1 + c_1/T)x_{1,t-1} + v_{1,t}$, and $x_{2,t} = (1 + c_2)x_{2,t-1} + v_{2,t}$ with

$$(v_{1,t}, v_{2,t}, u_t)^\top \sim iid\ \mathrm{N}\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix} 1 & -0.78 & 0.4 \\ -0.78 & 1 & 0.21 \\ 0.4 & 0.21 & 1 \end{pmatrix}\right).$$

For the convenience of comparison, we use the same covariance matrix of $(v_{1,t}, v_{2,t}, u_t)$ as Lee (2016). The persistency parameter c_1 is selected from $(0, -5, -10)$, while c_2 is selected from $(-0.1, -1.1)$. Obviously, $x_{1,t}$ is I1 or NI1 process, and $x_{2,t}$ is I0 process. The conditional quantile of y_t given $x_{1,t-1}$ and $x_{2,t-1}$ is given by

$$\begin{aligned} Q_{y_t}(\tau|\mathcal{F}_{t-1}) &= \mu[Q_{u_t}(\tau) + 3] + \beta_1[Q_{u_t}(\tau) + 3]x_{1t-1} + \beta_2[Q_{u_t}(\tau) + 3]x_{2t-1} \\ &= \mu_\tau + \beta_{1\tau}x_{1t-1} + \beta_{2\tau}x_{2t-1}, \end{aligned}$$

where $\mu_\tau = \mu[Q_{u_t}(\tau) + 3]$, $\beta_{1\tau} = \beta_1[Q_{u_t}(\tau) + 3]$, $\beta_{2\tau} = \beta_2[Q_{u_t}(\tau) + 3]$, and $Q_{u_t}(\tau)$ is the τth quantile of u_t.

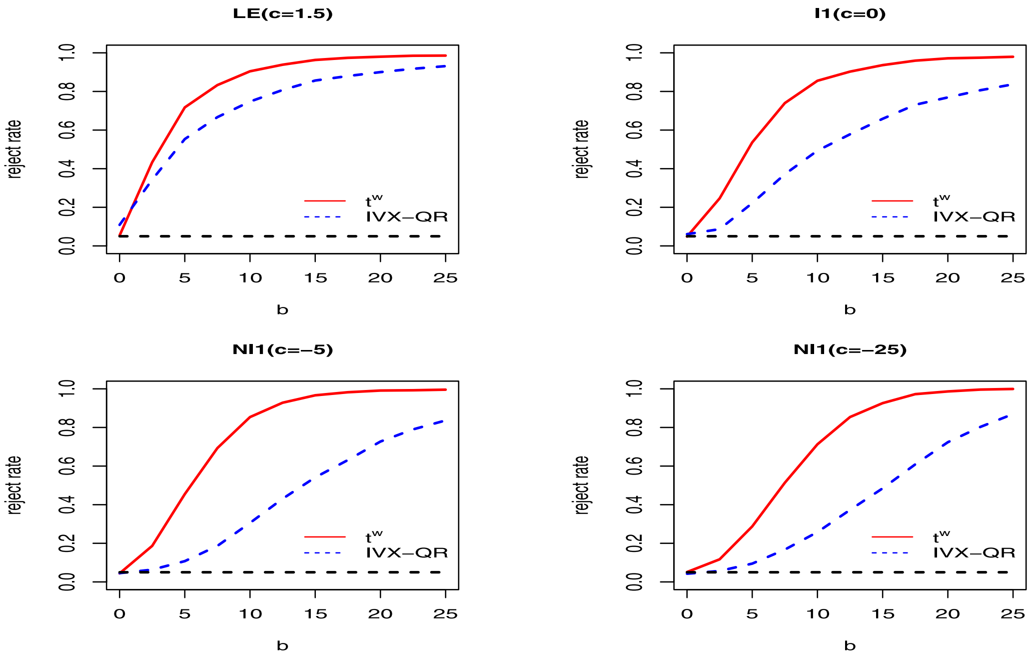

Fig. 1. Local power performances of t^w and IVX-QR for $\alpha = 1$, $\beta_\tau = b[Q_{u_t}(\tau) + 3]/T$, $\tau = 0.5$ and $T = 700$.

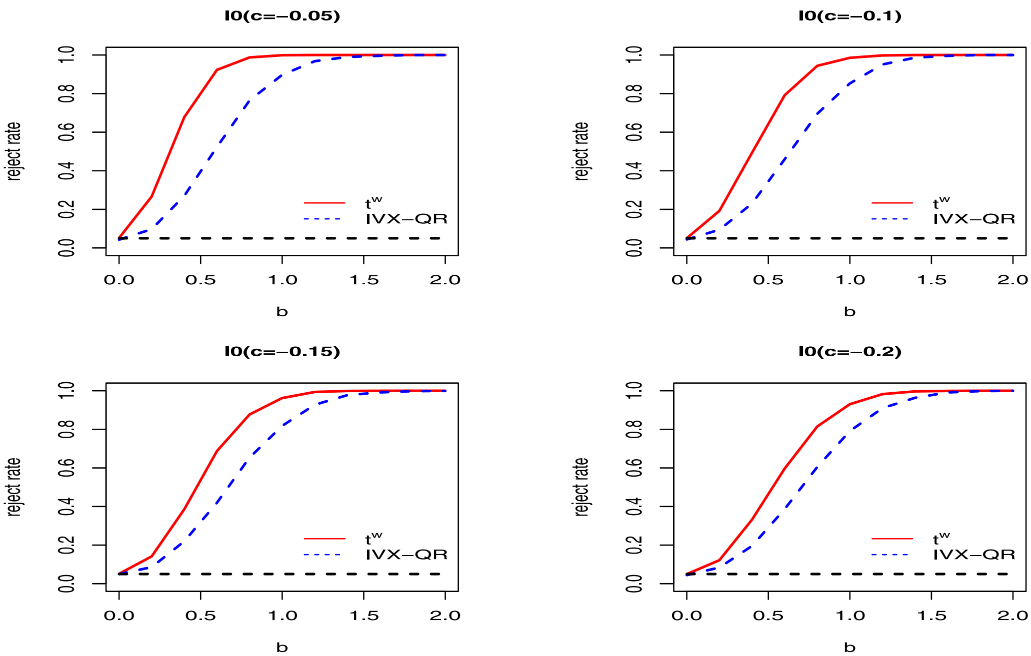

Fig. 2. Local power performances of t^w and IVX-QR for $\alpha = 0$, $\beta_\tau = b[Q_{u_t}(\tau) + 3]/\sqrt{T}$, $\tau = 0.5$ and $T = 700$.

First, we consider a joint hypothesis test: $H_0 : \beta_{1\tau} = \beta_{2\tau} = 0$. The sample size is set at $T = 700$ and the nominal size is defined as 5%. Similar to Example 1, the rejection rate is computed based on 500 simulations, and we repeated 100 times of simulations and computations for each setting. The mean and the standard error in parenthesis of 100 rejection rates are then reported given various combinations of persistency parameters (c_1, c_2) and at different quantile levels τ. Table 2 reports the size performance of both the IVX-QR approach and the proposed method. It can be observed that both

Table 2
Size performances of the joint test: $H_0 : \beta_{1\tau} = \beta_{2\tau} = 0$, with a nominal size of 5%.

Q_m^w											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$(c_1, c_2) = (0, -0.1)$	0.087	0.064	0.052	0.048	0.046	0.045	0.048	0.050	0.052	0.065	0.085
	(0.013)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.015)
$(c_1, c_2) = (-5, -0.1)$	0.086	0.065	0.051	0.049	0.046	0.046	0.047	0.047	0.053	0.067	0.085
	(0.012)	(0.011)	(0.010)	(0.010)	(0.008)	(0.008)	(0.009)	(0.010)	(0.010)	(0.012)	(0.012)
$(c_1, c_2) = (-10, -0.1)$	0.086	0.065	0.052	0.050	0.047	0.044	0.048	0.048	0.053	0.065	0.085
	(0.011)	(0.011)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.008)	(0.010)	(0.010)	(0.012)
$(c_1, c_2) = (0, -1.1)$	0.087	0.066	0.056	0.049	0.046	0.046	0.046	0.049	0.053	0.068	0.088
	(0.012)	(0.013)	(0.011)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.013)
$(c_1, c_2) = (-5, -1.1)$	0.087	0.065	0.052	0.050	0.048	0.046	0.047	0.050	0.053	0.068	0.086
	(0.011)	(0.010)	(0.010)	(0.010)	(0.008)	(0.009)	(0.011)	(0.010)	(0.009)	(0.012)	(0.012)
$(c_1, c_2) = (-10, -1.1)$	0.087	0.068	0.052	0.048	0.049	0.047	0.045	0.048	0.054	0.068	0.087
	(0.013)	(0.011)	(0.011)	(0.009)	(0.011)	(0.009)	(0.009)	(0.010)	(0.010)	(0.011)	(0.013)
IVX-QR											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$(c_1, c_2) = (0, -0.1)$	0.160	0.113	0.076	0.066	0.064	0.061	0.065	0.069	0.076	0.112	0.159
	(0.016)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)	(0.015)	(0.016)
$(c_1, c_2) = (-5, -0.1)$	0.095	0.070	0.053	0.049	0.046	0.046	0.047	0.049	0.054	0.069	0.097
	(0.013)	(0.012)	(0.010)	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.010)	(0.012)	(0.012)
$(c_1, c_2) = (-10, -0.1)$	0.091	0.069	0.052	0.047	0.045	0.045	0.045	0.047	0.051	0.068	0.092
	(0.013)	(0.012)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)	(0.011)	(0.011)	(0.012)	(0.013)
$(c_1, c_2) = (0, -1.1)$	0.149	0.106	0.073	0.061	0.058	0.057	0.056	0.062	0.071	0.103	0.150
	(0.016)	(0.014)	(0.011)	(0.010)	(0.010)	(0.011)	(0.011)	(0.010)	(0.012)	(0.013)	(0.018)
$(c_1, c_2) = (-5, -1.1)$	0.095	0.069	0.052	0.048	0.044	0.046	0.045	0.048	0.052	0.070	0.096
	(0.012)	(0.011)	(0.010)	(0.011)	(0.010)	(0.008)	(0.010)	(0.010)	(0.010)	(0.010)	(0.013)
$(c_1, c_2) = (-10, -1.1)$	0.087	0.067	0.052	0.047	0.044	0.044	0.045	0.046	0.051	0.067	0.090
	(0.014)	(0.012)	(0.010)	(0.010)	(0.009)	(0.008)	(0.009)	(0.009)	(0.011)	(0.012)	(0.011)

Note: The DGP is a bivariate quantile regression given by $y_t = 3(1 + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1}) + (1 + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1})u_t$ where $\beta_1 = \beta_2 = 0$, $x_{1,t} = (1 + c_1/T)x_{1,t-1} + v_{1,t}$, $x_{2,t} = (1 + c_2)x_{2,t-1} + v_{2,t}$.

approaches perform reasonably well under the joint null hypothesis. While the empirical sizes are closed to the nominal size 5% at inner quantiles, there still exist size distortions at extreme quantile levels for both methods. However, for most cases, the proposed method shows better size performance than the IVX-QR. Table 3 compares the power performances. It can be observed that two methods demonstrate comparable power performances. While our method has slightly better power performance at inner quantiles $\tau = 0.25, 0.5, 0.75$, the IVX-QR is more powerful at extreme quantiles $\tau = 0.05$ and 0.95.

Next, one may be interested in testing the predictability for each predictor in the multiple regression. Thus, we also conduct simulations for the single tests: $H_0 : \beta_{1\tau} = 0$ or $H_0 : \beta_{2\tau} = 0$, with the same DGP. Here, we only report the results of our method, as the IVX-QR approach based on the simple QR testing procedure is only consistent under the joint null hypothesis[9] as argued in Lee (2016). Table 4 reports the size performance of Q_m^w with a nominal size of 5%. From Panel A, which shows the results for the single test: $H_0 : \beta_{1\tau} = 0$, one may find that our method performs reasonably well, with a slight size distortions at extreme quantiles. Panel B reports the results for the single test: $H_0 : \beta_{2\tau} = 0$, and similar conclusions can be made.

Table 5 reports the power performance of Q_m^w with a nominal size of 5%. Panel A shows the results of the single test: $H_0 : \beta_{1\tau} = 0$, and it can be observed that the rejection rate converges to 1 as the value of β_1 increases, with a faster rate at the inner quantiles compared to extreme quantiles. Panel B displays the results of the single test: $H_0 : \beta_{2\tau} = 0$. Similarly, as the value of β_2 increases, the rejection rate converges to 1, demonstrating the effectiveness of our method. Comparing the results of both Panels A and B, it can be observed that the single test for the nonstationary predictor shows faster convergence rate than that for the stationary predictor, consistent with the findings by Cai and Wang (2014) on the different convergence rates for nonstationary and stationary regressors.

In summary, our method works reasonably well in terms of size and power for both univariate and bivariate predictive quantile models. Comparing to other existing methods in the literature, such as the VA approach and the Bonferroni Q-test, our method is still quite competitive.[10]

[9] Indeed, we also conduct simulations using the IVX-QR approach for the single tests, but it shows serious size distortions for most cases. The simulation results are not presented here since Lee (2016) did not consider this test.

[10] Following the suggestion from one of the referees, we extend the VA approach and the Bonferroni Q-test to predictive quantile regressions. The simulations, omitted here to save space, show that, for most cases, our method demonstrates comparable or even slightly better performance than these two methods. The codes and results are available upon request.

Table 3
Power performances of the joint test: $H_0 : \beta_{1\tau} = \beta_{2\tau} = 0$ with a nominal size of 5%.

	b_1	0	1.5	3	4.5	6	7.5	9	10.5	12	13.5	15
	b_2	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2
Q_m^w	$\tau = 0.05$	0.088	0.096	0.113	0.152	0.207	0.272	0.349	0.429	0.511	0.586	0.658
		(0.012)	(0.014)	(0.015)	(0.017)	(0.018)	(0.022)	(0.021)	(0.023)	(0.023)	(0.024)	(0.022)
	$\tau = 0.25$	0.051	0.087	0.203	0.381	0.565	0.730	0.848	0.913	0.953	0.974	0.985
		(0.009)	(0.012)	(0.017)	(0.021)	(0.023)	(0.021)	(0.016)	(0.011)	(0.009)	(0.007)	(0.006)
	$\tau = 0.5$	0.046	0.118	0.329	0.587	0.792	0.910	0.964	0.985	0.993	0.997	0.998
		(0.009)	(0.015)	(0.023)	(0.022)	(0.016)	(0.013)	(0.009)	(0.005)	(0.004)	(0.002)	(0.002)
	$\tau = 0.75$	0.051	0.145	0.408	0.688	0.867	0.952	0.983	0.993	0.997	0.999	0.999
		(0.010)	(0.017)	(0.024)	(0.019)	(0.015)	(0.009)	(0.006)	(0.003)	(0.002)	(0.002)	(0.001)
	$\tau = 0.95$	0.088	0.165	0.378	0.620	0.804	0.909	0.958	0.982	0.993	0.996	0.998
		(0.012)	(0.019)	(0.023)	(0.021)	(0.020)	(0.013)	(0.010)	(0.006)	(0.004)	(0.003)	(0.002)
IVX-QR	$\tau = 0.05$	0.156	0.157	0.189	0.246	0.324	0.423	0.515	0.600	0.680	0.743	0.800
		(0.016)	(0.019)	(0.017)	(0.019)	(0.021)	(0.023)	(0.023)	(0.022)	(0.021)	(0.020)	(0.020)
	$\tau = 0.25$	0.065	0.086	0.167	0.320	0.506	0.678	0.801	0.882	0.928	0.955	0.971
		(0.011)	(0.014)	(0.017)	(0.023)	(0.021)	(0.022)	(0.019)	(0.014)	(0.012)	(0.010)	(0.008)
	$\tau = 0.5$	0.053	0.092	0.227	0.458	0.688	0.849	0.930	0.968	0.986	0.993	0.996
		(0.010)	(0.012)	(0.019)	(0.024)	(0.020)	(0.014)	(0.011)	(0.008)	(0.005)	(0.004)	(0.003)
	$\tau = 0.75$	0.067	0.124	0.330	0.616	0.830	0.939	0.981	0.994	0.998	0.999	0.999
		(0.012)	(0.014)	(0.019)	(0.023)	(0.016)	(0.012)	(0.006)	(0.003)	(0.002)	(0.001)	(0.001)
	$\tau = 0.95$	0.157	0.260	0.544	0.795	0.927	0.979	0.995	0.999	1.000	1.000	1.000
		(0.014)	(0.022)	(0.021)	(0.017)	(0.011)	(0.006)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)

Note: The DGP is given by $y_t = 3(1+\beta_1 x_{1,t-1}+\beta_2 x_{2,t-1})+(1+\beta_1 x_{1,t-1}+\beta_2 x_{2,t-1})u_t$ where $\beta_1 = b_1/T$, $\beta_2 = b_2/\sqrt{T}$, $x_{1,t} = (1+c_1/T)x_{1,t-1}+v_{1,t}$, $x_{2,t} = (1+c_2)x_{2,t-1}+v_{2,t}$, with $c_1 = -1$ and $c_2 = -0.2$.

Table 4
Size performances of the single tests for Q_m^w, with a nominal size of 5% and $T = 700$.

Panel A: $H_0 : \beta_{1\tau} = 0$											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$(c_1, c_2) = (0, -0.1)$	0.080	0.065	0.052	0.051	0.048	0.047	0.047	0.050	0.052	0.062	0.078
	(0.014)	(0.011)	(0.011)	(0.010)	(0.012)	(0.010)	(0.009)	(0.010)	(0.010)	(0.011)	(0.013)
$(c_1, c_2) = (-5, -0.1)$	0.080	0.062	0.054	0.049	0.048	0.047	0.047	0.047	0.053	0.063	0.079
	(0.012)	(0.012)	(0.010)	(0.011)	(0.010)	(0.008)	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)
$(c_1, c_2) = (-10, -0.1)$	0.082	0.065	0.055	0.050	0.049	0.047	0.049	0.050	0.053	0.062	0.081
	(0.013)	(0.011)	(0.011)	(0.011)	(0.009)	(0.010)	(0.011)	(0.010)	(0.010)	(0.011)	(0.011)
$(c_1, c_2) = (0, -1.1)$	0.075	0.062	0.050	0.049	0.046	0.045	0.047	0.048	0.051	0.063	0.074
	(0.011)	(0.010)	(0.011)	(0.008)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)	(0.013)	(0.010)
$(c_1, c_2) = (-5, -1.1)$	0.076	0.062	0.052	0.048	0.047	0.047	0.048	0.049	0.054	0.062	0.073
	(0.012)	(0.011)	(0.011)	(0.010)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.012)	(0.012)
$(c_1, c_2) = (-10, -1.1)$	0.077	0.061	0.051	0.050	0.045	0.046	0.045	0.048	0.051	0.059	0.077
	(0.013)	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)	(0.010)	(0.012)
Panel B: $H_0 : \beta_{2\tau} = 0$											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$(c_1, c_2) = (0, -0.1)$	0.109	0.081	0.068	0.059	0.050	0.049	0.051	0.058	0.066	0.082	0.105
	(0.014)	(0.013)	(0.012)	(0.011)	(0.009)	(0.010)	(0.010)	(0.011)	(0.012)	(0.011)	(0.013)
$(c_1, c_2) = (-5, -0.1)$	0.080	0.067	0.055	0.050	0.049	0.050	0.049	0.050	0.052	0.065	0.080
	(0.013)	(0.012)	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)	(0.012)	(0.009)	(0.013)
$(c_1, c_2) = (-10, -0.1)$	0.080	0.065	0.054	0.051	0.048	0.047	0.049	0.050	0.054	0.063	0.077
	(0.013)	(0.012)	(0.010)	(0.010)	(0.009)	(0.008)	(0.009)	(0.009)	(0.011)	(0.012)	(0.009)
$(c_1, c_2) = (0, -1.1)$	0.088	0.065	0.052	0.047	0.045	0.042	0.044	0.048	0.050	0.065	0.088
	(0.012)	(0.010)	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.010)	(0.009)	(0.011)	(0.013)
$(c_1, c_2) = (-5, -1.1)$	0.081	0.064	0.051	0.049	0.048	0.047	0.047	0.049	0.054	0.064	0.081
	(0.012)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)	(0.009)	(0.011)	(0.012)
$(c_1, c_2) = (-10, -1.1)$	0.082	0.063	0.052	0.050	0.048	0.047	0.048	0.048	0.052	0.064	0.081
	(0.011)	(0.011)	(0.010)	(0.010)	(0.009)	(0.010)	(0.009)	(0.010)	(0.010)	(0.011)	(0.012)

Note: The DGP is given by $y_t = 3(1+\beta_1 x_{1,t-1}+\beta_2 x_{2,t-1})+(1+\beta_1 x_{1,t-1}+\beta_2 x_{2,t-1})u_t$ where $x_{1,t} = (1+c_1/T)x_{1,t-1}+v_{1,t}$, $x_{2,t} = (1+c_2)x_{2,t-1}+v_{2,t}$. For Panel A, $\beta_1 = 0$ and $\beta_2 = 2/\sqrt{T}$. For Panel B, $\beta_1 = 15/T$ and $\beta_2 = 0$.

Table 5
Power performances of the single tests for Q_m^w, with a nominal size of 5% and $T = 700$.

Panel A: $H_0: \beta_{1\tau} = 0$											
b_1	0	4	8	12	16	20	24	28	32	36	40
$\tau = 0.05$	0.076	0.119	0.245	0.401	0.551	0.666	0.753	0.816	0.855	0.881	0.904
	(0.013)	(0.014)	(0.018)	(0.021)	(0.021)	(0.021)	(0.019)	(0.018)	(0.014)	(0.014)	(0.013)
$\tau = 0.25$	0.051	0.270	0.605	0.807	0.899	0.942	0.964	0.973	0.979	0.982	0.985
	(0.010)	(0.022)	(0.021)	(0.019)	(0.014)	(0.009)	(0.008)	(0.006)	(0.006)	(0.005)	(0.005)
$\tau = 0.5$	0.045	0.405	0.765	0.907	0.957	0.976	0.983	0.987	0.990	0.991	0.993
	(0.009)	(0.024)	(0.020)	(0.011)	(0.010)	(0.007)	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)
$\tau = 0.75$	0.050	0.474	0.825	0.935	0.970	0.982	0.988	0.990	0.993	0.994	0.994
	(0.010)	(0.024)	(0.018)	(0.011)	(0.008)	(0.005)	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)
$\tau = 0.95$	0.075	0.435	0.777	0.910	0.958	0.976	0.982	0.987	0.988	0.990	0.989
	(0.012)	(0.022)	(0.017)	(0.013)	(0.009)	(0.007)	(0.006)	(0.005)	(0.004)	(0.005)	(0.004)
Panel B: $H_0: \beta_{2\tau} = 0$											
b_2	0	0.3	0.6	0.9	1.2	1.5	1.8	2.1	2.4	2.7	3
$\tau = 0.05$	0.087	0.101	0.148	0.218	0.307	0.407	0.511	0.614	0.703	0.776	0.838
	(0.012)	(0.014)	(0.015)	(0.019)	(0.021)	(0.023)	(0.023)	(0.023)	(0.018)	(0.018)	(0.016)
$\tau = 0.25$	0.052	0.135	0.351	0.604	0.804	0.910	0.962	0.986	0.993	0.997	0.998
	(0.010)	(0.017)	(0.019)	(0.022)	(0.018)	(0.014)	(0.009)	(0.005)	(0.003)	(0.003)	(0.002)
$\tau = 0.5$	0.048	0.201	0.544	0.821	0.943	0.983	0.994	0.998	0.998	0.999	0.999
	(0.010)	(0.018)	(0.022)	(0.016)	(0.010)	(0.006)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
$\tau = 0.75$	0.051	0.250	0.638	0.886	0.971	0.992	0.997	0.999	0.999	0.999	0.999
	(0.010)	(0.019)	(0.024)	(0.015)	(0.008)	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
$\tau = 0.95$	0.085	0.249	0.571	0.820	0.938	0.980	0.993	0.997	0.998	0.999	0.999
	(0.013)	(0.020)	(0.023)	(0.016)	(0.011)	(0.006)	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)

Note: The DGP is given by $y_t = 3(1+\beta_1 x_{1,t-1} + \beta_2 x_{2,t-1}) + (1+\beta_1 x_{1,t-1} + \beta_2 x_{2,t-1})u_t$ where $x_{1,t} = (1+c_1/T)x_{1,t-1} + v_{1,t}$, $x_{2,t} = (1+c_2)x_{2,t-1} + v_{2,t}$ with $c_1 = -1$ and $c_2 = -0.2$. For Panel A, $\beta_1 = b_1/T$ and $\beta_2 = 2/\sqrt{T}$. For Panel B, $\beta_1 = 15/T$ and $\beta_2 = b_2/\sqrt{T}$.

Table 6
95% confidence intervals for ρ in different sample periods.

Predictor	1927–2002	1927–2005	1927–2018	1952–2002	1952–2005	1952–2018
d/p	[0.983, 1.000]	[0.985, 1.000]	[0.986, 1.000]	[0.988, 1.003]	[0.989, 1.002]	[0.989, 1.002]
e/p	[0.979, 0.999]	[0.978, 0.997]	[0.978, 0.996]	[0.986, 1.003]	[0.984, 1.001]	[0.980, 0.999]
b/m	[0.971, 0.994]	[0.973, 0.995]	[0.976, 0.995]	[0.985, 1.001]	[0.985, 1.001]	[0.987, 1.001]
ntis	[0.957, 0.987]	[0.957, 0.987]	[0.971, 0.993]	[0.954, 0.990]	[0.954, 0.989]	[0.970, 0.995]
d/e	[0.991, 1.001]	[0.993, 1.002]	[0.983, 0.998]	[0.989, 1.001]	[0.993, 1.003]	[0.975, 0.997]
tbl	[0.984, 0.999]	[0.984, 0.999]	[0.986, 0.999]	[0.976, 1.000]	[0.976, 0.999]	[0.982, 0.999]
dfy	[0.962, 0.989]	[0.962, 0.989]	[0.961, 0.987]	[0.954, 0.990]	[0.954, 0.989]	[0.953, 0.986]
tms	[0.936, 0.974]	[0.938, 0.975]	[0.925, 0.964]	[0.921, 0.972]	[0.926, 0.973]	[0.914, 0.961]

5. An empirical application

5.1. Data

This section applies the newly proposed method to re-investigate whether or not stock market index returns are predictable by a set of macroeconomic indicators and financial ratios. For a convenient comparison, our main results are based on the same data set (monthly data) in Lee (2016), with a sample period from January 1927 to December 2005. An updated data set until December 2018 is considered too to see whether there is any change after the 2008 global crisis.[11] The dependent variable is stock market excess returns, which is computed as the difference between S&P 500 index (including dividends) monthly returns and the one-month Treasury bill rate. Following the literature, eight popular predictors are considered, including dividend–price (d/p), earnings–price (e/p), book to market ratios (b/m), net equity expansion (*ntis*), dividend–payout ratio (d/e), T-bill rate (*tbl*), default yield spread (*dfy*), term spread (*tms*).[12] These predictors are standard in the predictive regression literature, and could be further classified into three categories: valuation ratios (d/p, e/p and b/m), corporate finance variables (*ntis* and d/e) and bond yield measures (*tbl*, *tms* and *dfy*), see Cenesizoglu and Timmermann (2008) and Lee (2016).

Table 6 reports the 95% confidence interval of the first-order autocorrelation coefficient ρ for the eight predicting variables during different sample periods. All predictors show strong evidence of high persistency for all periods, but we

11 The updated data set can be downloaded from http://www.hec.unil.ch/agoyal, the website of Professor Amit Goyal.

12 One may refer to Welch and Goyal (2008) for details on how to construct economic foundations of all variables.

Table 7
P-values (%) of quantile prediction tests using the univariate model (1927:01–2005:12).

τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	**0.2**	**0.3**	**2.7**	22.1	23.2	25.2	13.5	**1.3**	**0.3**	**0.5**	**0.2**
e/p	**0.1**	**0.7**	6.0	25.9	26.0	22.9	17.2	**3.6**	**1.0**	**0.8**	**1.0**
b/m	8.4	13.5	40.6	44.8	50.1	68.2	18.7	**4.2**	**0.6**	**0.3**	**0.0**
ntis	**2.9**	**0.4**	**0.1**	9.9	14.1	10.0	45.8	56.5	64.8	58.1	57.3
d/e	**0.0**	**0.0**	**0.0**	**0.3**	8.8	45.2	33.8	**2.3**	**0.0**	**0.0**	**0.0**
tbl	7.7	10.1	41.3	27.1	**2.3**	**0.8**	5.9	7.0	**0.6**	**1.0**	**2.1**
dfy	**0.0**	**0.0**	**0.0**	**0.0**	**4.5**	57.3	**0.7**	**0.0**	**0.0**	**0.0**	**0.0**
tms	36.8	42.6	33.4	61.4	35.6	49.9	79.5	66.0	54.1	15.5	13.3

Note: p-values are in bold if less than or equal to the significant level 5%.

Table 8
P-values (%) of quantile prediction tests using the univariate model (1952:01–2005:12).

τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	18.6	28.6	36.7	27.2	22.8	31.1	36.6	28.0	39.7	10.4	9.5
e/p	18.6	30.7	40.1	27.4	20.2	31.6	39.0	30.1	36.0	16.2	12.3
b/m	53.7	64.2	64.6	59.6	16.4	32.8	72.5	43.8	48.1	50.7	22.1
ntis	39.4	38.7	18.6	41.5	27.6	23.1	24.7	6.0	**4.3**	**0.7**	**0.5**
d/e	12.2	31.0	43.4	17.2	12.3	46.9	48.6	68.1	47.3	13.6	28.3
tbl	5.8	**3.4**	**2.5**	**1.1**	**0.4**	**2.6**	22.0	72.3	71.9	40.4	4.3
dfy	50.9	71.6	65.9	46.0	30.2	70.5	22.9	17.3	**2.9**	**0.4**	**0.0**
tms	8.5	**1.7**	6.7	26.6	17.7	36.8	77.5	44.9	76.6	52.0	48.2

Note: p-values are in bold if less than or equal to the significant level 5%.

are still unable to identify the persistency category for each variable, see Fan and Lee (2019). Given that our new method is robust to all persistency categories, it is expected to provide more reliable conclusions than traditional approaches developed under a specific type of persistency.

5.2. Empirical results

First, we investigate the quantile predictability of stock returns for each individual predictor using the univariate model, and then analyze the predictability of individual predictor and different combinations of predictors in the framework of multivariate quantile regressions.

Table 7 reports the univariate regression results given the sample period from January 1927 to December 2005. The p-values (%) shown in bold imply the rejection of the null hypothesis of no predictability at the 5% level. The main findings can be summarized as follows. For the group of valuation ratios, we find significant lower and upper quantiles predictability for both d/p and e/p ratios, but only upper quantiles predictability for the b/m ratio. For the group of corporate finance predictors, the d/e, which represents the corporation dividend payment policy, has strong predictability at both lower and upper quantiles, while the *ntis*, measuring the corporate issuing activity, has predictive ability at lower quantiles only. For the group of bond yield measures, the *dfy* shows significant predictability at most quantiles except at median level, and the *tbl* is significant at upper quantiles. However, we do not find any evidence of the significant predictability for the *tms* at all quantiles. Compared to Lee (2016), we obtain similar testing results for d/p, d/e, *ntis*, *tbl* and *dfy*, but different results for the other three. For the b/m ratio, Lee (2016) found significant predictability for both lower and upper quantiles, while only upper quantiles predictability for our method. For the e/p, we find both lower and upper quantiles predictability, but Lee (2016) only reported a significant predictability at the 80% quantile level. Meanwhile, Lee (2016) found significant predictability at upper quantiles (0.9 and 0.95) for the *tms*, for which we do not find any significant predictive ability. The difference is reasonable as our method corrects the size distortion and enjoys improve the power due to a faster convergence rate of the estimator, compared to IVX-QR approach. Meanwhile, our testing results show smoother changes across different quantiles, demonstrating a better performance on robustness and stability.

Next, we conduct the quantile prediction tests for the post-1952 data until December 2005, and report the results in Table 8. Compared with Table 7, in general, there are fewer variables with significant predicting power, implying that the market efficiency is improved after World War II, see Campbell and Yogo (2006). Especially, we do not find any significant predictability for value ratios (d/p, e/p and b/m) and the d/e ratio. For lower quantiles, only *tbl* and *tms* still have significant predictive ability, while *ntis* and *dfy* are significant for upper quantiles. For middle quantiles, only *tbl* has significant predicting power. Compared to Lee (2016), we share a similar finding that the bond yield measures, especially the *tbl* and *dfy*, maintain the significant quantile predictability, but we find a weaker predicting power of value ratios during the sample period 1952:01–2005:12.

Because the stock returns might be affected by multiple variables, the univariate model may exaggerate the prediction power for each variable. Therefore, we re-examine the stock market predictability in the framework of multivariate

Table 9
P-values (%) for the test using the multivariate model (1927:01–2005:12).

Ang and Bekaert (2007)											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	**0.9**	**0.8**	**4.2**	11.9	24.0	33.6	22.7	7.3	**3.2**	**0.6**	**1.0**
tbl	**0.1**	**0.1**	**1.3**	9.0	18.9	20.8	19.6	10.9	**1.9**	**0.1**	**0.4**
Joint test	**0.1**	**0.0**	**1.0**	10.9	26.4	17.3	5.5	**0.6**	**0.1**	**0.0**	**0.1**
Ferson and Schadt (1996)											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	**0.1**	**0.2**	**2.7**	**1.1**	7.9	9.0	27.0	18.9	26.1	14.5	9.2
tbl	**0.3**	**0.4**	**3.0**	**3.7**	11.4	11.4	14.5	8.7	**2.3**	**0.7**	**0.3**
dfy	**0.0**	**0.0**	**0.3**	**0.4**	5.7	11.2	28.0	17.1	**1.6**	**0.1**	**0.0**
tms	**0.0**	**0.1**	**0.6**	**0.6**	4.7	4.9	24.8	20.6	31.0	16.2	8.4
Joint test	**0.0**	**0.0**	**0.0**	**0.6**	11.7	7.2	**2.1**	**0.0**	**0.0**	**0.0**	**0.0**
Kothari and Shanken (1997)											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	**0.1**	**0.1**	**0.7**	13.9	15.4	17.2	15.8	**3.3**	**0.5**	**0.1**	**0.0**
b/m	**0.2**	**0.1**	**1.3**	10.8	17.3	20.2	24.4	8.7	**1.1**	**0.3**	**0.0**
Joint test	**0.0**	**0.0**	**1.0**	16.9	22.5	24.6	14.2	**0.7**	**0.1**	**0.1**	**0.0**
Lamont (1998)											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	5.2	**3.7**	8.8	18.0	17.0	25.4	28.8	12.9	9.5	**2.0**	**1.0**
d/e	**0.4**	**0.1**	**0.7**	9.2	18.5	22.7	25.3	10.7	**1.5**	**0.1**	**0.0**
Joint test	**0.2**	**0.0**	**0.8**	19.2	20.0	28.2	17.0	**0.7**	**0.1**	**0.0**	**0.0**
Campbell and Vuolteenaho (2004)											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
e/p	**0.0**	**0.0**	**0.3**	10.1	15.7	19.7	13.7	**1.8**	**0.2**	**0.0**	**0.0**
b/m	**0.0**	**0.0**	**0.1**	7.7	13.7	19.3	15.2	**2.3**	**0.2**	**0.0**	**0.0**
tms	**0.0**	**0.1**	**0.7**	30.1	24.2	26.8	15.9	**2.6**	**0.2**	**0.1**	**0.0**
Joint test	**0.0**	**0.0**	**0.0**	6.5	14.0	36.6	13.4	**0.7**	**0.0**	**0.0**	**0.0**
Full Model											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	7.9	14.9	16.1	14.2	17.5	18.6	21.7	20.8	10.1	**2.3**	**2.1**
e/p	6.2	**5.0**	**1.5**	**1.0**	**1.4**	**4.6**	7.2	10.7	13.0	6.5	5.3
b/m	9.8	18.1	20.6	13.2	16.5	15.6	17.1	16.3	8.8	**2.4**	**2.1**
ntis	**1.3**	5.3	8.8	**3.5**	11.8	11.9	24.5	16.7	7.5	**1.8**	**1.6**
tbl	12.6	24.3	22.0	10.0	7.5	7.3	6.3	**4.3**	**3.1**	**1.5**	**1.6**
dfy	**0.0**	**0.0**	**1.0**	6.3	27.3	21.9	**2.8**	**0.1**	**0.0**	**0.0**	**0.0**
tms	**3.0**	10.6	14.8	13.7	19.4	18.4	18.9	17.5	9.3	**1.9**	**1.1**
Joint test	**0.0**	**0.0**	**0.0**	**0.0**	**0.1**	**0.8**	**0.1**	**0.0**	**0.0**	**0.0**	**0.0**

Note: *p*-values are in bold if less than or equal to the significant level 5%. For the full model, d/e is excluded due to the multiple collinearity among d/e, d/p and e/p ratios.

predictive quantile regression. Following Kostakis et al. (2015), we consider five popular prediction models in the literature and a full model with seven predictors (d/e is excluded due to the multiple collinearity). For each model, we report the single test results for each individual predictor and the joint test results for the combination of all predictors.

Table 9 depicts the test results during the sample period from January 1927 to December 2005. Interestingly, both single tests and joint tests based on the first five predictive models do not find any significant predictability at middle quantile levels, confirming the existing findings about a weak predictability at the mean/median of stock returns. However, all five models show evidence of significant predictability at lower and upper quantiles, suggesting a stronger predictability in the extreme market status. For the full model, after controlling other variables, some predictors lose their prediction power, though the joint tests suggest that the full model has prediction power at all quantiles. It worths to be mentioned that the bond yield measures, including *tbl*, *dfy* and *tms*, maintain the significant predictability at either lower quantiles or upper quantiles or both. The persistency of the predictive ability for these macroeconomic variables is further confirmed in Table 10, where only the predictive models containing bond yield measures keep prediction power in the post-1952 sample period. Because Lee (2016) only considered a bivariate case, a comparison with the results is not provided here.

To see whether there is any change on the market predictability in the recent years, we apply our method to the most updated data set for two sample periods: 1927:01-2018:12 and the post-1952 sample period 1952:01-2018:12. The main conclusions are roughly consistent with those using the sample period until December 2005.[13]

[13] To save space, we skip to report the empirical results for the updated data set, which are available upon request.

Table 10
P-values (%) for the test using the multivariate model (1952:01–2005:12).

Ang and Bekaert (2007)											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	22.3	32.0	40.6	27.8	35.5	33.7	29.1	23.2	33.1	9.5	9.0
tbl	13.4	31.1	21.4	15.3	13.1	17.2	22.5	24.2	28.1	10.7	**2.4**
Joint test	18.0	37.5	35.1	21.0	19.9	26.8	32.0	26.8	42.7	11.2	**2.1**
Ferson and Schadt (1996)											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	18.6	17.3	11.1	7.6	9.7	12.8	9.6	7.2	5.8	**2.1**	7.0
tbl	21.0	15.8	7.4	5.6	7.6	11.2	5.4	**4.5**	6.2	**2.4**	**5.0**
dfy	18.9	17.0	7.5	5.5	8.3	10.4	7.3	**5.0**	5.2	**2.0**	**4.7**
tms	23.0	16.6	8.9	7.0	7.5	12.2	6.9	5.3	6.3	**2.2**	6.3
Joint test	24.7	27.7	23.2	13.5	16.9	29.9	15.2	11.7	6.4	**0.4**	**0.7**
Kothari and Shanken (1997)											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	13.3	18.4	26.1	17.0	14.0	20.1	21.0	23.4	25.5	10.3	5.8
b/m	12.9	20.3	23.2	16.3	12.1	18.4	20.3	23.1	24.7	8.6	5.2
Joint test	16.8	27.6	36.2	25.2	16.2	23.2	28.7	26.6	36.1	11.9	5.2
Lamont (1998)											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	22.9	25.8	31.6	30.5	23.1	22.2	28.2	23.3	29.1	18.5	11.7
d/e	16.2	21.9	23.7	17.5	13.1	20.7	26.6	23.6	25.5	10.8	5.1
Joint test	20.2	28.4	41.1	27.6	20.4	33.0	38.6	29.0	39.3	9.7	6.2
Campbell and Vuolteenaho (2004)											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
e/p	14.3	15.1	17.1	12.8	10.7	14.9	18.1	16.8	23.0	11.3	11.7
b/m	10.1	22.4	14.9	9.3	11.0	14.8	17.6	16.9	24.8	14.8	15.2
tms	17.6	29.8	28.4	19.1	16.5	29.0	27.3	27.7	27.6	7.1	6.9
Joint test	18.5	37.4	35.7	18.4	13.0	29.9	37.5	27.9	30.7	5.3	**1.5**
Full Model											
τ	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	16.3	19.6	18.7	13.7	11.5	11.7	15.6	23.4	27.0	16.4	10.9
e/p	14.6	18.4	17.6	19.8	18.9	25.2	21.6	16.3	15.0	7.8	**4.4**
b/m	11.5	18.7	14.7	13.2	15.7	15.2	14.7	13.4	15.5	8.9	7.9
ntis	19.4	18.1	8.8	12.9	14.7	19.9	16.5	17.8	16.0	7.2	10.5
tbl	**1.5**	**4.8**	8.5	**4.9**	**4.2**	**4.5**	**5.0**	**4.9**	6.7	5.8	**4.9**
dfy	7.2	15.2	19.8	16.7	13.4	12.5	9.4	5.6	5.3	**1.7**	**2.4**
tms	17.7	16.4	16.2	21.6	19.1	23.2	22.8	22.3	21.7	11.5	8.3
Joint test	**0.1**	**0.6**	**0.3**	**0.4**	**0.3**	**1.1**	**1.8**	**1.1**	**1.2**	**0.0**	**0.0**

Note: *p*-values are in bold if less than or equal to the significant level 5%. For the full model, d/e is excluded due to the multiple collinearity among d/e, d/p and e/p ratios.

6. Conclusion

This paper investigates the inferential theory for predictive quantile regression with highly persistent predictors, containing both the stationary case and the nonstationary case. A weighted estimator based on the VA approach is proposed to construct the pivotal test statistic. By introducing a new additional variable whose key component is independent of x_t in NI1, I1 and LE cases and persistency is the same as that for x_t, our method is not only free of the size distortion but it can also achieve the local power under the optimal rate T with nonstationary predictors and $\sqrt{T}$ with stationary predictors. The numerical performance of the proposed tests is checked by simulation studies which show that the proposed method outperforms the IVX-QR approach proposed by Lee (2016) in a finite sample. In the empirical application, we apply the new method to test the predictability of US stock returns at different quantile levels. Interestingly, after the World War II, we do not find much evidence for the prediction power for some well-known financial ratios, such as e/p ratio, d/p ratio and b/m ratio. However, the macroeconomic indicators, such as dfy, tms and tbl show strong evidence of significant prediction power, especially at lower and upper quantile levels.

Finally, we note several possible extensions of the present study. For example, it may be interesting to extend the model in (1) to a nonparametric form similar to the model in Cai and Xu (2008) for stationary case. Also, one might extend our model to the case that $0 < \alpha < 1$, corresponding to the so-called mildly integrated MI processes ($c < 0$) or mildly explosive ME processes ($c > 0$), as mentioned in Section 2.1. We leave such extensions as possible future research topics.

Acknowledgments

The authors thank the Editor (Professor Xiaohong Chen) and three anonymous reviewers for their helpful and constructive comments to improve significantly the quality of the paper. Also, we thank Professor Ji-Hyung Lee for sharing the empirical data set and code for IVX-QR. This work, in part, was supported by the National Natural Science Foundation of China with Grant Numbers #71631004, #72033008, #71850011, #71571152 and #71988101, and the Fundamental Research Funds for the Central Universities with Grant Number #JBK2001034, as well as the fund provided by Fujian Provincial Key Laboratory of Statistics (Xiamen University) with grant number #2019001. We also thank the valuable comments from the participants of "The Third Forum of Chinese Econometricians" in Dalian and "The 2019 Guangzhou Econometrics Workshop".

Appendix A. Mathematical proofs

Note that due to the limitation of space, the brief derivations of the main results are only provided and all lemmas with the detailed proof are presented in the online appendix, which can be found at http://www.people.ku.edu/~z397c158/CCL-Supplement.pdf.

Proof of Theorem 1. As the proof for I0 case is standard, we only provide the proof for NI1, I1 and LE cases. Define the convex object function $Z_T(v) = \sum_{t=2}^{T}\left\{\rho_\tau\left[u_{t\tau} - v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}\right] - \rho_\tau(u_{t\tau})\right\}$. Following Xiao (2009), the minimization problem in (5) is equivalent to minimizing $Z_T(v)$, i.e., if $\hat{v}$ is the minimizer of $Z_T(v)$, then $\hat{v} = \mathbf{D}_T(\hat{\theta}_\tau - \theta_\tau)$, where $\hat{\theta}_\tau$ is the solution of the minimization problem in (5).

Now, an application of the Knight identity as in Knight (1989) implies that

$$Z_T(v) = -\sum_{t=2}^{T} v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}\psi_\tau(u_{t\tau}) + \sum_{t=2}^{T}\int_0^{v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}} \left[1(u_{t\tau} \le l) - 1(u_{t\tau} \le 0)\right] dl.$$

Note that $Z_T(v)$ is derivable with respect with v, and define the new objective function $V_T(v) = -\partial Z_T(v)/\partial v$. It is easy to show that

$$V_T(v) = \sum_{t=2}^{T} \mathbf{D}_T^{-1}\Lambda_{t-1}\psi_\tau(u_{t\tau}) - \sum_{t=2}^{T} \mathbf{D}_T^{-1}\Lambda_{t-1}\left[1(u_{t\tau} \le v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right]. \tag{A.1}$$

To prove Theorem 1, it suffices to verify that $V_T(v)$ satisfies the conditions of Lemma A.1 in the online appendix. We first check the condition (i), i.e., $-v^\top V_T(\lambda v) \ge -v^\top V_T(v)$ for $\lambda \ge 1$. Note that

$$-v^\top V_T(\lambda v) = -v^\top \sum_{t=2}^{T} \mathbf{D}_T^{-1}\Lambda_{t-1}\psi_\tau(u_{t\tau}) + \sum_{t=2}^{T} v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}\left[1(u_{t\tau} \le \lambda v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right],$$

so that it needs to show that $-v^\top V_T(\lambda v)$ is a non-decreasing function of λ. Given the fact that $1(u \le x) - 1(u < 0)$ is a non-decreasing function of x, one can show that $1(u_{t\tau} \le \lambda v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)$ is a non-decreasing function of λ if $v^\top \mathbf{D}_T^{-1}\Lambda_{t-1} > 0$, and a decreasing function of λ if $v^\top \mathbf{D}_T^{-1}\Lambda_{t-1} < 0$. Thus, $-v^\top V_T(\lambda v)$ is always a non-decreasing function of λ, and the condition (i) in Lemma A.1 is verified.

Next, we check the condition (ii), i.e., $\sup_{\|v\|<M}\|V_T(v) + f_{u_\tau}(0)Nv - A_T\| = o_p(1)$ for $0 < M < \infty$. Define $\eta_t = \mathbf{D}_T^{-1}\Lambda_{t-1}\left[1(u_{t\tau} \le v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right]$ and $A_T = \sum_{t=2}^{T}\mathbf{D}_T^{-1}\Lambda_{t-1}\psi_\tau(u_{t\tau})$. Clearly, it follows from (A.1) that

$$V_T(v) = A_T - \sum_{t=2}^{T} E_{t-1}(\eta_t) - \sum_{t=2}^{T}\left[\eta_t - E_{t-1}(\eta_t)\right]. \tag{A.2}$$

Therefore, to prove the condition (ii), it suffices to show that $\sum_{t=2}^{T} E_{t-1}(\eta_t) = f_{u_\tau}(0)Nv + o_p(1)$ and $\sum_{t=2}^{T}\left[\eta_t - E_{t-1}(\eta_t)\right] = o_p(1)$. By Taylor expansion,

$$\begin{aligned}\sum_{t=2}^{T} E_{t-1}(\eta_t) &= \sum_{t=2}^{T}\mathbf{D}_T^{-1}\Lambda_{t-1}\left[F_{u_{t\tau},t-1}(v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) - F_{u_{t\tau},t-1}(0)\right]\\ &= \sum_{t=2}^{T}\mathbf{D}_T^{-1}\Lambda_{t-1}\left[f_{u_{t\tau},t-1}(0)\Lambda_{t-1}^\top \mathbf{D}_T^{-1}v + \frac{1}{2}f'_{u_{t\tau},t-1}(l^*)v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \mathbf{D}_T^{-1}v\right]\\ &= B_1 + B_2,\end{aligned}$$

where $l^* \in (0, v^\top \mathbf{D}_T^{-1}\Lambda_{t-1})$ if $v^\top \mathbf{D}_T^{-1}\Lambda_{t-1} > 0$ and $l^* \in (v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}, 0)$ if $v^\top \mathbf{D}_T^{-1}\Lambda_{t-1} < 0$. By Assumption A.2 and stationarity of $f_{u_{t\tau},t-1}(0)$, $\sup_{0\le r\le 1}\left|\frac{1}{T^{1-\delta}}\sum_{t=2}^{\lfloor rT\rfloor}\left[f_{u_{t\tau},t-1}(0) - f_{u_\tau}(0)\right]\right| = o_p(1)$ for some $\delta > 0$ (Xiao, 2009). Also, by Lemma

A.2 in the online appendix, $N_T = \mathbf{D}_T^{-1}\sum_{t=2}^T \Lambda_{t-1}\Lambda_{t-1}^\top \mathbf{D}_T^{-1} = N + o_p(1)$, where N is positive definite random matrix defined in Lemma A.2. By using the similar idea as in Xiao (2009), it is easy to show that

$$B_1 = \sum_{t=2}^T \left[f_{u_{t\tau},t-1}(0) - f_{u_\tau}(0)\right]\mathbf{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \mathbf{D}_T^{-1} v + \sum_{t=2}^T f_{u_\tau}(0)\mathbf{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \mathbf{D}_T^{-1} v$$
$$= f_{u_\tau}(0)Nv + o_p(1).$$

To prove $B_2 = o_p(1)$, first,

$$\|B_2\| = \left\| \sum_{t=2}^T \mathbf{D}_T^{-1}\Lambda_{t-1}\frac{1}{2}f'_{u_{t\tau},t-1}(l^*)v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \mathbf{D}_T^{-1} v\right\|$$
$$\le \frac{1}{\sqrt{T}}\frac{1}{2}\sup_{x\in\mathbb{R}}|f'_{u_{t\tau},t-1}(x)|\sum_{t=2}^T \sqrt{T}\|\mathbf{D}_T^{-1}\Lambda_{t-1}\|v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \mathbf{D}_T^{-1} v.$$

By Assumption A.2, $\sup_{x\in\mathbb{R}}|f'_{u_{t\tau},t-1}(x)| = O_p(1)$, and by Lemma A.3 in the online appendix, $\sum_{t=2}^T \sqrt{T}\|\mathbf{D}_T^{-1}\Lambda_{t-1}\|v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \mathbf{D}_T^{-1} v = O_p(1)$ for any $\|v\| < M$, $0 < M < \infty$. Then, $\|B_2\| \le \frac{1}{\sqrt{T}}O_p(1)O_p(1) = o_p(1)$. Combining the above results, we have

$$\sum_{t=2}^T E_{t-1}(\eta_t) = f_{u_\tau}(0)Nv + o_p(1). \tag{A.3}$$

Furthermore, it is to verify the fact that $\sum_{t=2}^T[\eta_t - E_{t-1}(\eta_t)] = o_p(1)$. Note that

$$\sum_{t=2}^T[\eta_t - E_{t-1}(\eta_t)] = \begin{pmatrix} \sum_{t=2}^T[\eta_{1t} - E_{t-1}(\eta_{1t})] \\ \sum_{t=2}^T[\eta_{2t} - E_{t-1}(\eta_{2t})] \\ \sum_{t=2}^T[\eta_{3t} - E_{t-1}(\eta_{3t})]\end{pmatrix},$$

where $\eta_{1t} = \left[1(u_{t\tau} \le v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right]/\sqrt{T}$, $\eta_{2t} = x_{t-1}^*\left[1(u_{t\tau} \le v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right]/T$, and $\eta_{3t} = \left[1(u_{t\tau} \le v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right]/T$. To save space, we provide the detailed proof for η_{1t} as an illustration to show that $\sum_{t=2}^T[\eta_{it} - E_{t-1}(\eta_{it})] = o_p(1)$, $i = 1, 2$, and 3 and skip the rest for η_{2t} and η_{3t}. For some $2 \le t \le T$ satisfying $v^\top \mathbf{D}_T^{-1}\Lambda_{t-1} > 0$, $1(u_{t\tau} \le v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0) = 1(0 < u_{t\tau} \le v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) \in [0, 1]$, one can show that

$$\begin{aligned} T\cdot E_{t-1}(\eta_{1t}^2) &\le E_{t-1}\left[1(u_{t\tau} \le v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right] = F_{u_{t\tau}}(v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}|\mathcal{F}_{t-1}) - F_{u_{t\tau}}(0|\mathcal{F}_{t-1}) \\ &= \underset{l_t\in(0,v^\top \mathbf{D}_T^{-1}\Lambda_{t-1})}{f_{u_{t\tau},t-1}(l_t)}\; v^\top \mathbf{D}_T^{-1}\Lambda_{t-1} \le \sup_{x\in\mathbb{R}}|f_{u_{t\tau},t-1}(x)|\cdot|v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}|. \end{aligned} \tag{A.4}$$

The last step holds by Taylor expansion. Similarly, for any $2 \le t \le T$ satisfying $v^\top \mathbf{D}_T^{-1}\Lambda_{t-1} \le 0$, $1(u_{t\tau} \le 0) - 1(u_{t\tau} \le v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) = 1(v^\top \mathbf{D}_T^{-1}\Lambda_{t-1} < u_{t\tau} \le 0) \in [0, 1]$, it is easy to verify that

$$T\cdot E_{t-1}(\eta_{1t}^2) = E_{t-1}\left[-1(u_{t\tau} \le v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}) + 1(u_{t\tau} \le 0)\right]^2 \le \sup_{x\in\mathbb{R}}|f_{u_{t\tau},t-1}(x)||v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}|. \tag{A.5}$$

Then, it follows by (A.4) and (A.5) that

$$E_{t-1}(\eta_{1t}^2) \le \frac{1}{T}\sup_{x\in\mathbb{R}}|f_{u_{t\tau},t-1}(x)|\cdot|v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}|.$$

Therefore,

$$\sum_{t=2}^T E_{t-1}(\eta_{1t}^2) \le \frac{1}{T}\sum_{t=2}^T \sup_{x\in\mathbb{R}}|f_{u_{t\tau},t-1}(x)|\cdot|v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}| \le \sup_{x\in\mathbb{R}}|f_{u_{t\tau},t-1}(x)|\frac{1}{T}\sum_{t=2}^T|v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}|,$$

Note that $\sup_{x\in\mathbb{R}}|f'_{u_{t\tau},t-1}(x)| = O_p(1)$ by Assumption A.2, and $\frac{1}{\sqrt{T}}\sum_{t=2}^T|v^\top \mathbf{D}_T^{-1}\Lambda_{t-1}| = O_p(1)$, for any $\|v\| < M$, $0 < M < \infty$. Thus,

$$\sum_{t=2}^T E_{t-1}(\eta_{1t}^2) \le O_p(1)O_p(1/\sqrt{T}) = o_p(1).$$

By the same token, one can obtain that $\sum_{t=2}^T[E_{t-1}(\eta_{1t})]^2 = o_p(1)$ and the fact that $[\eta_{1t} - E_{t-1}(\eta_{1t})]$ is MDS implies that $Var\left(\sum_{t=2}^T[\eta_{1t} - E_{t-1}(\eta_{1t})]\right) = o_p(1)$. Hence, $\sum_{t=2}^T[\eta_{1t} - E_{t-1}(\eta_{1t})] = o_p(1)$. Similarly, one can show that

$\sum_{t=2}^{T}[\eta_{2t}-E_{t-1}(\eta_{2t})]=o_p(1)$ and $\sum_{t=2}^{T}[\eta_{3t}-E_{t-1}(\eta_{3t})]=o_p(1)$. Therefore,

$$\sum_{t=2}^{T}[\eta_t-E_{t-1}(\eta_t)]=o_p(1). \tag{A.6}$$

By (A.2), (A.3) and (A.6), for any $\|v\|<M$, $0<M<\infty$, $V_T(v)=A_T-f_{u_\tau}(0)Nv+o_p(1)$. Therefore, for $0<M<\infty$, $\sup_{\|v\|<M}\|V_T(v)+f_{u_\tau}(0)Nv-A_T\|=o_p(1)$ holds. By Lemma A.2, it is straightforward to show that $\|A_T\|=O_p(1)$.

Clearly, all conditions of Lemma A.1 are verified so that an application of Lemma A.1 leads to

$$\mathbf{D}_T(\hat{\theta}_\tau-\theta_\tau)=f_{u_\tau}(0)^{-1}N_T^{-1}\mathbf{D}_T^{-1}\sum_{t=2}^{T}\Lambda_{t-1}\psi_\tau(u_{t\tau})+o_p(1),$$

which completes the proof of Theorem 1. □

Proof of Theorem 2. For simplicity, we only offer the proof for the NI1, I1 and LE cases, because the proof for the case I0 case is standard. By the Bahadur representation stated in Theorem 1, we have

$$\begin{pmatrix} 1 & \sum_{t=2}^{T}\frac{x_{t-1}^*}{T^{3/2}} & \sum_{t=2}^{T}\frac{z_{t-1}}{T^{3/2}} \\ \sum_{t=2}^{T}\frac{x_{t-1}^*}{T^{3/2}} & \sum_{t=2}^{T}\frac{(x_{t-1}^*)^2}{T^2} & \sum_{t=2}^{T}\frac{z_{t-1}x_{t-1}^*}{T^2} \\ \sum_{t=2}^{T}\frac{z_{t-1}}{T^{3/2}} & \sum_{t=2}^{T}\frac{x_{t-1}^*z_{t-1}}{T^2} & \sum_{t=2}^{T}\frac{z_{t-1}^2}{T^2} \end{pmatrix}\begin{pmatrix}\sqrt{T}(\hat{\mu}_\tau-\mu_\tau)\\ T(\hat{\beta}_\tau-\beta_\tau)\\ T(\hat{\gamma}_\tau-\beta_\tau)\end{pmatrix}=f_{u_\tau}(0)^{-1}\begin{pmatrix}\sum_{t=2}^{T}\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}}\\ \sum_{t=2}^{T}\frac{x_{t-1}^*}{\sqrt{T}}\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}}\\ \sum_{t=2}^{T}\frac{z_{t-1}}{\sqrt{T}}\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}}\end{pmatrix}+o_p(1). \tag{A.7}$$

Define $S\equiv\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ -\frac{1}{T^{3/2}}\sum_{t=2}^{T}z_{t-1} & 0 & 1\end{pmatrix}$. Then, by pre-multiplying S on both sides of (A.7), one has,

$$\begin{pmatrix} 1 & \sum_{t=2}^{T}\frac{x_{t-1}^*}{T^{3/2}} & \sum_{t=2}^{T}\frac{z_{t-1}}{T^{3/2}} \\ \sum_{t=2}^{T}\frac{x_{t-1}^*}{T^{3/2}} & \sum_{t=2}^{T}\frac{(x_{t-1}^*)^2}{T^2} & \sum_{t=2}^{T}\frac{z_{t-1}x_{t-1}^*}{T^2} \\ 0 & W_1 & W_2 \end{pmatrix}\begin{pmatrix}\sqrt{T}(\hat{\mu}_\tau-\mu_\tau)\\ T(\hat{\beta}_\tau-\beta_\tau)\\ T(\hat{\gamma}_\tau-\beta_\tau)\end{pmatrix}=f_{u_\tau}(0)^{-1}\begin{pmatrix}\sum_{t=2}^{T}\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}}\\ \sum_{t=2}^{T}\frac{x_{t-1}^*\psi_\tau(u_{t\tau})}{T}\\ \sum_{t=2}^{T}\frac{\bar{z}_{t-1}\psi_\tau(u_{t\tau})}{T}\end{pmatrix}+o_p(1),$$

where $W_1=\sum_{t=2}^{T}x_{t-1}^*z_{t-1}/T^2-\sum_{t=2}^{T}z_{t-1}\sum_{t=2}^{T}x_{t-1}^*/T^3$, $W_2=\sum_{t=2}^{T}z_{t-1}^2/T^2-\left(\sum_{t=2}^{T}z_{t-1}\right)^2/T^3$, and $\bar{z}_{t-1}=z_{t-1}-\sum_{t=2}^{T}z_{t-1}/T$. From the third row in the above equation, we have

$$(W_1+W_2)T(\hat{\beta}_\tau^w-\beta_\tau)=f_{u_\tau}(0)^{-1}\sum_{t=2}^{T}\left(\frac{z_{t-1}}{\sqrt{T}}-\frac{1}{T^{3/2}}\sum_{t=2}^{T}z_{t-1}\right)\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}}+o_p(1). \tag{A.8}$$

Recall the definition $z_{t-1}=\hat{\pi}_1\zeta_{t-1}+x_{t-1}/\sqrt{1+x_{t-1}^2}$, and for the NI1, I1 and LE cases, $\zeta_{\lfloor rT\rfloor}/\sqrt{T}\Rightarrow B(r)$, $\hat{\pi}_1\xrightarrow{d}\tilde{\pi}_1$ and $x_{t-1}/\sqrt{1+x_{t-1}^2}=O_p(1)$. Thus, $z_{\lfloor rT\rfloor}/\sqrt{T}\Rightarrow\tilde{\pi}_1B(r)$. By the continuous mapping theorem, one obtains that

$$(W_1+W_2)T(\hat{\beta}_\tau^w-\beta_\tau)\xrightarrow{d}f_{u_\tau}(0)^{-1}\tilde{\pi}_1\int\left[B(r)-\int B(r)dr\right]_{\perp}dB_{\psi_\tau}(r).$$

From (11), $W_1+W_2\xrightarrow{d}\tilde{\pi}_1\int\bar{B}(r)\bar{J}_x^c(r)dr$. Using the independence between ζ_t and $u_{t\tau}$, we have

$$T(\hat{\beta}_\tau^w-\beta_\tau)\xrightarrow{d}f_{u_\tau}(0)^{-1}\,\mathrm{MN}\left[0,\,\tau(1-\tau)\frac{\int\bar{B}(r)^2dr}{[\int\bar{B}(r)\bar{J}_x^c(r)dr]^2}\right].$$

This ends the proof of the theorem. □

Proof of Theorem 3. We only offer the proof for NI1, I1 and LE case since the proof for the I0 case is standard. For NI1, I1 and LE case, $z_{\lfloor rT \rfloor}/\sqrt{T} \Rightarrow \tilde{\pi}_1 B(r)$. It follows that

$$W_2 = \sum_{t=2}^{T} z_{t-1}^2/T^2 - \left(\sum_{t=2}^{T} z_{t-1}\right)^2/T^3 = \sum_{t=2}^{T}\left(z_{t-1} - \frac{1}{T}\sum_{t=2}^{T} z_{t-1}\right)^2/T^2 \xrightarrow{d} \int \bar{B}(r)^2 dr.$$

By the continuous mapping theorem and Slutsky Theorem,

$$\begin{aligned} t^w &= \hat{f}_{u_\tau}(0)\left[W_2\tau(1-\tau)\right]^{-1/2}(W_1+W_2)T(\hat{\beta}_\tau^w - \beta_\tau) \\ &\xrightarrow{d} f_{u_\tau}(0)\left[\tau(1-\tau)\int \bar{B}(r)^2 dr\right]^{-1/2} f_{u_\tau}(0)^{-1}\,\mathrm{MN}\left(0, \tau(1-\tau)\int \bar{B}(r)^2 dr\right) \overset{d}{=} \mathrm{N}(0,1). \end{aligned}$$

Moreover, under the local alternative hypothesis $H_a : \beta_\tau = \frac{b_\tau}{T}$, it follows that

$$\begin{aligned} &\hat{f}_{u_\tau}(0)\left[W_2\tau(1-\tau)\right]^{-1/2}(W_1+W_2)T\beta_\tau \\ &= \hat{f}_{u_\tau}(0)\left[W_2\tau(1-\tau)\right]^{-1/2}(W_1+W_2)b_\tau \xrightarrow{d} b_\tau \frac{f_{u_\tau}(0)}{\sqrt{\tau(1-\tau)}} \frac{\tilde{\pi}_1 \int \bar{B}(r)J_x^c(r)dr}{\sqrt{\tilde{\pi}_1^2 \int \bar{B}(r)^2 dr}} \\ &= b_\tau \frac{f_{u_\tau}(0)}{\sqrt{\tau(1-\tau)}} \frac{\tilde{\pi}_1 \int \bar{B}(r)J_x^c(r)dr}{|\tilde{\pi}_1|\sqrt{\int \bar{B}(r)^2 dr}} = b_\tau \frac{f_{u_\tau}(0)}{\sqrt{\tau(1-\tau)}} \frac{\mathrm{sign}(\tilde{\pi}_1)\int \bar{B}(r)J_x^c(r)dr}{\sqrt{\int \bar{B}(r)^2 dr}} \\ &= b_\tau \frac{f_{u_\tau}(0)}{\sqrt{\tau(1-\tau)}} \frac{\mathrm{sign}(\tilde{\pi}_1)\,\mathrm{sign}(\tilde{\pi}_1)\left|\int \bar{B}(r)J_x^c(r)dr\right|}{\sqrt{\int \bar{B}(r)^2 dr}} \\ &= b_\tau \frac{f_{u_\tau}(0)}{\sqrt{\tau(1-\tau)}} \frac{\mathrm{sign}(\tilde{\pi}_1)^2\left|\int \bar{B}(r)J_x^c(r)dr\right|}{\sqrt{\int \bar{B}(r)^2 dr}} = b_\tau|\pi_c|/\sigma_\tau. \end{aligned}$$

Therefore,

$$\begin{aligned} t^w &= \hat{f}_{u_\tau}(0)\left[W_2\tau(1-\tau)\right]^{-1/2}(W_1+W_2)T\hat{\beta}_\tau^w \\ &= \hat{f}_{u_\tau}(0)\left[W_2\tau(1-\tau)\right]^{-1/2}(W_1+W_2)T(\hat{\beta}_\tau^w - \beta_\tau) + \hat{f}_{u_\tau}(0)\left[W_2\tau(1-\tau)\right]^{-1/2}(W_1+W_2)T\beta_\tau \\ &\xrightarrow{d} b_\tau|\pi_c|/\sigma_\tau + B(1). \end{aligned}$$

This concludes the proof the theorem. □

Proof of Theorem 4. Similar to the proof of the Bahadur representation theorem for the univariate case, one can establish the Bahadur representation for multivariate quantile regressions. To save a space, the details are omitted. Now,

$$\begin{aligned} \tilde{D}_T(\hat{\boldsymbol{\beta}}_\tau^w - \boldsymbol{\beta}_\tau) =& f_{u_\tau}(0)^{-1}\left[(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)\mathbf{X}_{t-1}^\top(\tilde{D}_T)^{-1}\right]^{-1} \\ &\cdot(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)\psi_\tau(u_{t\tau}) + o_p(1). \end{aligned} \tag{A.9}$$

Note that for all predictors $x_{i,t}$, $i = 1, 2$,

$$\begin{cases} z_{i,\lfloor rT \rfloor}/\sqrt{T} = \hat{\pi}_{1,i}\zeta_{1,t-1}\left[1 + o_p(1)\right], & \text{if } x_{i,t} \text{ is NI1, I1 and LE;} \\ z_{i,t} = x_{i,t}/\sqrt{1 + x_{i,t}^2} + o_p(1), & \text{if } x_{i,t} \text{ is I0.} \end{cases}$$

First, we consider Case 1, $K_1 = 0$, i.e., all predictors are stationary. Then,

$$z_t = (z_{1,t}, z_{2,t})^\top = \left(x_{1,t}/\sqrt{1 + x_{1,t}^2},\, x_{2,t}/\sqrt{1 + x_{2,t}^2}\right)^\top + o_p(1),$$

and the weighting matrix $\tilde{D}_T = \mathrm{diag}(\sqrt{T}, \sqrt{T})$. By the central limit theorem, it is easy to show that

$$(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)\psi_\tau(u_{t\tau}) \xrightarrow{d} \mathrm{N}(0, \tau(1-\tau)\mathbf{V}_2), \tag{A.10}$$

where

$$\mathbf{V}_2 = \text{var}\left[\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)\right] = \begin{pmatrix} \text{Var}\left(\frac{x_{1,t}}{\sqrt{1+x_{1,t}^2}}\right) & \text{Cov}\left(\frac{x_{1,t}}{\sqrt{1+x_{1,t}^2}}, \frac{x_{2,t}}{\sqrt{1+x_{2t}^2}}\right) \\ \text{Cov}\left(\frac{x_{1,t}}{\sqrt{1+x_{1,t}^2}}, \frac{x_{2,t}}{\sqrt{1+x_{2t}^2}}\right) & \text{Var}\left(\frac{x_{2,t}}{\sqrt{1+x_{2,t}^2}}\right) \end{pmatrix}.$$

A Combination of (A.9) and (A.10), together with the continuous mapping theorem, leads to

$$\tilde{D}_T(\hat{\boldsymbol{\beta}}_\tau^w - \boldsymbol{\beta}_\tau) \xrightarrow{d} f_{u_\tau}(0)^{-1}\mathbf{V}_1^{-1}\,\text{N}\left(0, \tau(1-\tau)\mathbf{V}_2\right),$$

where

$$\mathbf{V}_1 = \operatorname*{plim}_{T\to\infty}(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)\mathbf{X}_{t-1}^\top(\tilde{D}_T)^{-1} = \begin{pmatrix} E\left(\frac{x_{1,t}^2}{\sqrt{1+x_{1,t}^2}}\right) & E\left(\frac{x_{1,t}x_{2,t}}{\sqrt{1+x_{1,t}^2}}\right) \\ E\left(\frac{x_{1,t}x_{2,t}}{\sqrt{1+x_{2,t}^2}}\right) & E\left(\frac{x_{2,t}^2}{\sqrt{1+x_{2,t}^2}}\right) \end{pmatrix}.$$

Next, for Case 2, $K_1 = 1$, i.e., x_{1t} is nonstationary and x_{2t} is stationary. Then,

$$\sqrt{T}\left(\tilde{D}_T\right)^{-1}\mathbf{Z}_t = (z_{1,t}/\sqrt{T}, z_{2,t})^\top = \left(\hat{\pi}_{1,1}\zeta_{1,t-1}/\sqrt{T}, x_{2,t-1}/\sqrt{1+x_{2,t-1}^2}\right)^\top + o_p(1),$$

and the weighting matrix $\tilde{D}_T = \text{diag}(T, \sqrt{T})$. Define $\bar{G} = \text{diag}\left(\sqrt{\frac{1}{T^2}\sum_{t=2}^{T}\bar{\zeta}_{1,t-1}^2}, 1\right)$ and

$$h_{t-1} = \left(h_{1,t-1}, h_{2,t-1}\right)^\top = (\tilde{D}_T)^{-1}\left[\frac{\bar{\zeta}_{1,t-1}}{\sqrt{\frac{1}{T^2}\sum_{t=2}^{T}\bar{\zeta}_{1,t-1}^2}}, \frac{x_{2,t-1}}{\sqrt{1+x_{2,t-1}^2}} - \frac{1}{T}\sum_{t=2}^{T}\frac{x_{2,t-1}}{\sqrt{1+x_{2,t-1}^2}}\right]^\top.$$

Thus,

$$\bar{G}^{-1}(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)\psi_\tau(u_{t\tau}) = \sum_{t=2}^{T}h_{t-1}\psi_\tau(u_{t\tau}) + o_p(1). \tag{A.11}$$

Thus, by Lemma A.4 in the online appendix, the Lindeberg condition for $h_{t-1}\psi_\tau(u_{t\tau})$ holds. That is, for any $\tilde{\varepsilon} > 0$,

$$\sum_{t=2}^{T}E\left[\|h_{t-1}\psi_\tau(u_{t\tau})\|^2\ 1\left(\|h_{t-1}\psi_\tau(u_{t\tau})\| > \tilde{\varepsilon}\right)\middle|\,\mathcal{F}_{t-1}\right] \xrightarrow{p} 0. \tag{A.12}$$

Again, by Lemma A.5 in the online appendix, the asymptotic variance of $\sum_{t=2}^{T}h_{t-1}\psi_\tau(u_{t\tau})$ is given by

$$\begin{aligned}\sum_{t=2}^{T}E\left[h_{t-1}h_{t-1}^\top\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}\right] &= \sum_{t=2}^{T}E\left[\begin{pmatrix} h_{1,t-1}^2 & h_{1,t-1}h_{2,t-1} \\ h_{1,t-1}h_{2,t-1} & h_{2,t-1}^2\end{pmatrix}\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}\right] \\ &\xrightarrow{p} \tau(1-\tau)\begin{pmatrix} 1 & 0 \\ 0 & \text{Var}\left(x_{2,t-1}/\sqrt{1+x_{2,t-1}}\right)\end{pmatrix}.\end{aligned} \tag{A.13}$$

Moreover, it is straightforward that $[h_{t-1}\psi_\tau(u_{t\tau})]_{t=2}^T$ is martingale difference sequence. Therefore, it follows by (A.12) and (A.13) and the Corollary 3.1 in Hall and Heyde (1980) that

$$\sum_{t=2}^{T}h_{t-1}\psi_\tau(u_{t\tau}) \xrightarrow{d} \text{N}\left[0, \tau(1-\tau)\begin{pmatrix} 1 & 0 \\ 0 & \text{Var}\left(x_{2,t-1}/\sqrt{1+x_{2,t-1}}\right)\end{pmatrix}\right]. \tag{A.14}$$

It is easy to see by (A.11) and (A.14) that

$$(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)\psi_\tau(u_{t\tau}) \xrightarrow{d} \text{MN}\left(0, \tau(1-\tau)\mathbf{V}_2\right), \tag{A.15}$$

where $\mathbf{V}_2 = \begin{pmatrix} \tilde{\pi}_{1,1}^2 \int \bar{B}_1(r)^2 \, dr & 0 \\ 0 & E\left(x_{2,t}^2/\left(1+x_{2,t}^2\right)\right) - E\left(x_{2,t}/\sqrt{1+x_{2,t}^2}\right)^2 \end{pmatrix}$. Next, an application of (A.9) and (A.15) as well as the continuous mapping theorem implies that

$$\tilde{D}_T(\hat{\boldsymbol{\beta}}_\tau^w - \boldsymbol{\beta}_\tau) \xrightarrow{d} f_{u_\tau}(0)^{-1}\mathbf{V}_1^{-1}\, \mathrm{MN}\left(0, \tau(1-\tau)\mathbf{V}_2\right),$$

where

$$\begin{aligned}
\mathbf{V}_1 &= \operatorname*{plim}_{T\to\infty}(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)\mathbf{X}_{t-1}^\top(\tilde{D}_T)^{-1} \\
&= \operatorname*{plim}_{T\to\infty}(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\left(\mathbf{X}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{X}_{t-1}\right)^\top(\tilde{D}_T)^{-1} \\
&= \operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\left(\hat{\pi}_{1,1}\frac{\zeta_{1,t-1}}{\sqrt{T}}, \frac{x_{2,t-1}}{\sqrt{1+x_{2,t-1}^2}}\right)^\top\left(\frac{x_{1,t-1}}{\sqrt{T}} - \frac{1}{T}\sum_{t=2}^{T}\frac{x_{1,t-1}}{\sqrt{T}}, x_{2,t-1} - \frac{1}{T}\sum_{t=2}^{T}x_{2,t-1}\right) \\
&= \begin{pmatrix} \tilde{\pi}_{1,1}\int \bar{B}_1(r)J_{x_1}^{c_1}(r)dr & 0 \\ 0 & E\left(x_{2,t}^2/\sqrt{1+x_{2,t}^2}\right) \end{pmatrix}.
\end{aligned}$$

Finally, for Case 3, $K_1 = 2$, i.e., all predictors are nonstationary, it is clear to see that

$$\sqrt{T}\left(\tilde{D}_T\right)^{-1}\mathbf{Z}_t = (z_{1,t}/\sqrt{T}, z_{2,t}/\sqrt{T})^\top = \left(\hat{\pi}_{1,1}\zeta_{1,t-1}/\sqrt{T}, \hat{\pi}_{1,2}\zeta_{2,t-1}/\sqrt{T}\right)^\top + o_p(1),$$

and the weighting matrix $\tilde{D}_T = \mathrm{diag}(T, T)$. Similar to the univariate model, one can show easily that

$$\begin{aligned}
(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)\psi_\tau(u_{t\tau}) &\xrightarrow{d} \int\left(\tilde{\pi}_{1,1}\bar{B}_1(r), \tilde{\pi}_{1,2}\bar{B}_2(r)\right)_\perp^\top dB_{\psi_\tau}(r) \\
&= \mathrm{MN}\left(0, \tau(1-\tau)\mathbf{V}_2\right),
\end{aligned} \tag{A.16}$$

where

$$\begin{aligned}
\mathbf{V}_2 &= \operatorname*{plim}_{T\to\infty}(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)^\top(\tilde{D}_T)^{-1} \\
&= \begin{pmatrix} \tilde{\pi}_{1,1}^2\int \bar{B}_1(r)^2 dr & \tilde{\pi}_{1,1}\tilde{\pi}_{1,2}\int \bar{B}_1(r)\bar{B}_2(r)dr \\ \tilde{\pi}_{1,1}\tilde{\pi}_{1,2}\int \bar{B}_2(r)\bar{B}_1(r)dr & \tilde{\pi}_{1,2}^2\int \bar{B}_2(r)^2 dr \end{pmatrix}.
\end{aligned}$$

The asymptotic mixture normality holds by the independence between $(\zeta_{1,t}, \zeta_{2,t})^\top$ and $\psi_\tau(u_{t\tau})$. Again, it follows by combining (A.9) and (A.16) together with the continuous mapping theorem that $\tilde{D}_T(\hat{\boldsymbol{\beta}}_\tau^w - \boldsymbol{\beta}_\tau) \xrightarrow{d} f_{u_\tau}(0)^{-1}\mathbf{V}_1^{-1}\,\mathrm{MN}\left(0, \tau(1-\tau)\mathbf{V}_2\right)$, where

$$\begin{aligned}
\mathbf{V}_1 &= \operatorname*{plim}_{T\to\infty}(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\left(\mathbf{Z}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\right)\mathbf{X}_{t-1}^\top(\tilde{D}_T)^{-1} \\
&= \operatorname*{plim}_{T\to\infty}(\tilde{D}_T)^{-1}\sum_{t=2}^{T}\mathbf{Z}_{t-1}\left(\mathbf{X}_{t-1} - \frac{1}{T}\sum_{t=2}^{T}\mathbf{X}_{t-1}\right)^\top(\tilde{D}_T)^{-1} \\
&= \operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\left(\tilde{\pi}_{1,1}\frac{\zeta_{1,t-1}}{\sqrt{T}}, \tilde{\pi}_{1,2}\frac{\zeta_{2,t-1}}{\sqrt{T}}\right)^\top\left(\frac{x_{1,t-1}}{\sqrt{T}} - \frac{1}{T}\sum_{t=2}^{T}\frac{x_{1,t-1}}{\sqrt{T}}, \frac{x_{2,t-1}}{\sqrt{T}} - \frac{1}{T}\sum_{t=2}^{T}\frac{x_{2,t-1}}{\sqrt{T}}\right) \\
&= \begin{pmatrix} \tilde{\pi}_{1,1}\int \bar{B}_1(r)J_{x_1}^{c_1}(r)dr & \tilde{\pi}_{1,1}\int \bar{B}_1(r)J_{x_2}^{c_2}(r)dr \\ \tilde{\pi}_{1,2}\int \bar{B}_2(r)J_{x_1}^{c_1}(r)dr & \tilde{\pi}_{1,2}\int \bar{B}_2(r)J_{x_2}^{c_2}(r)dr \end{pmatrix}.
\end{aligned}$$

This concludes the proof the theorem. □

Proof of Theorem 5. By the results in Theorem 4, the proof of Theorem 5 is straightforward and the details are omitted here to save space. □

Appendix B. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2021.10.012. The supplementary material with some necessary lemmas and their proofs to support the main theorems in the paper can be found online at http://www.people.ku.edu/~z397c158/CCL-Supplement.pdf.

References

Amihud, Y., Hurvich, C., Wang, Y., 2009. Multiple-predictor regressions: Hypothesis testing. Rev. Financ. Stud. 22 (1), 413–434.
Ang, A., Bekaert, G., 2007. Stock return predictability: Is it there? Rev. Financ. Stud. 20 (3), 651–707.
Bauer, D., Maynard, A., 2012. Persistence-robust surplus-lag Granger causality testing. J. Econometrics 169 (2), 293–300.
Breitung, J., Demetrescu, M., 2015. Instrumental variable and variable addition based inference in predictive regressions. J. Econometrics 187 (1), 358–375.
Cai, Z., Wang, Y., 2014. Testing predictive regression models with nonstationary regressors. J. Econometrics 178 (1), 4–14.
Cai, Z., Xu, X., 2008. Nonparametric quantile estimations for dynamic smooth coefficient models. J. Amer. Statist. Assoc. 103 (484), 1595–1608.
Campbell, J.Y., Vuolteenaho, T., 2004. Bad beta, good beta. Amer. Econ. Rev. 94 (5), 1249–1275.
Campbell, J.Y., Yogo, M., 2006. Efficient tests of stock return predictability. J. Financ. Econ. 81 (1), 27–60.
Cenesizoglu, T., Timmermann, A.G., 2008. Is the distribution of stock returns predictable. Available at SSRN 1107185.
Choi, Y., Jacewitz, S., Park, J.Y., 2016. A reexamination of stock return predictability. J. Econometrics 192 (1), 168–189.
Demetrescu, M., Rodrigues, P.M.M., 2020. Residual-augmented IVX predictive regression. J. Econometrics http://dx.doi.org/10.1016/j.jeconom.2020.11.007.
Dittmar, R.F., 2002. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. J. Financ. 57 (1), 369–403.
Dolado, J.J., Lütkepohl, H., 1996. Making wald tests work for cointegrated VAR systems. Econom. Rev. 15 (4), 369–386.
Elliott, G., 2011. A control function approach for testing the usefulness of trending variables in forecast models and linear regression. J. Econometrics 164 (1), 79–91.
Fan, R., Lee, J.H., 2019. Predictive quantile regressions under persistence and conditional heteroskedasticity. J. Econometrics 213 (1), 261–280.
Ferson, W.E., Schadt, R.W., 1996. Measuring fund strategy and performance in changing economic conditions. J. Financ. 51 (2), 425–461.
Gaglianone, W.P., Lima, L.R., Linton, O., Smith, D.R., 2011. Evaluating value-at-risk models via quantile regression. J. Bus. Econom. Statist. 29 (1), 150–160.
Gonzalo, J., Pitarakis, J.Y., 2012. Regime-specific predictability in predictive regressions. J. Bus. Econom. Statist. 30 (2), 229–241.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press, New York.
Harvey, C.R., Siddique, A., 2000. Conditional skewness in asset pricing tests. J. Financ. 55 (3), 1263–1295.
Hosseinkouchack, M., Demetrescu, M., 2021. Finite-sample size control of IVX-based tests in predictive regressions. Econom. Theory 37 (4), 769–793.
Jansson, M., Moreira, M.J., 2006. Optimal inference in regression models with nearly integrated regressors. Econometrica 74 (3), 681–714.
Knight, K., 1989. Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. Canad. J. Statist. 17 (3), 261–278.
Koenker, R., 2005. Quantile Regression. Cambridge Press, New York.
Kostakis, A., Magdalinos, T., Stamatogiannis, M.P., 2015. Robust econometric inference for stock return predictability. Rev. Financ. Stud. 28 (5), 1506–1553.
Kothari, S.P., Shanken, J., 1997. Book-to-market, dividend yield, and expected market returns: A time series analysis. J. Financ. Econ. 44 (2), 169–203.
Lamont, O.A., 1998. Earnings and expected returns. J. Financ. 53 (5), 1563–1587.
Lee, J.H., 2016. Predictive quantile regression with persistent covariates: IVX-QR approach. J. Econometrics 192 (1), 105–118.
Liao, X., Cai, Z., Chen, H., 2018. A perspective on recent models for testing predictability of asset returns. Appl. Math. J. Chin. Univ. Ser. B 33 (2), 127–144.
Liu, X., Yang, B., Cai, Z., Peng, L., 2019. A unified test for predictability of asset returns regardless of properties of predicting variables. J. Econometrics 208 (1), 141–159.
Magdalinos, T., Phillips, P.C.B., 2009. Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. Econom. Theory 25 (2), 482–526.
Maynard, A., Shimotsu, K., Wang, Y., 2011. Inference in Predictive Quantile Regressions. Working Paper, Department of Economics, University of Guelph.
Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74 (3), 535–547.
Phillips, P.C.B., 1989. Partially identified econometric models. Econom. Theory 5 (2), 181–240.
Phillips, P.C.B., 2014. Optimal estimation of cointegrated systems with irrelevant instruments. J. Econometrics 178 (2), 210–224.
Phillips, P.C.B., 2015. Halbert White Jr. memorial JFEC lecture: Pitfalls and possibilities in predictive regression. J. Financ. Econ. 13 (3), 521–555.
Phillips, P.C.B., Lee, J.H., 2013. Predictive regression under various degrees of persistence and robust long-horizon regression. J. Econometrics 177 (2), 250–264.
Phillips, P.C.B., Lee, J.H., 2016. Robust econometric inference with mixed integrated and mildly explosive regressors. J. Econometrics 192 (2), 433–450.
Phillips, P.C.B., Magdalinos, T., 2009. Econometric Inference in the Vicinity of Unity. CoFieWorking Paper No. 7, Singapore Management University.
Phillips, P.C.B., Shi, S., Yu, J., 2015. Testing for multiple bubbles: Limit theory of real time detectors. Internat. Econom. Rev. 56 (4), 1079–1133.
Toda, H.Y., Yamamoto, T., 1995. Statistical inference in vector autoregressions with possibly integrated processes. J. Econometrics 66 (1), 225–250.
Torous, W., Valkanov, R., Yan, S., 2004. On predicting stock returns with nearly integrated explanatory variables. J. Bus. 77 (4), 937–966.
Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21 (4), 1455–1508.
Xiao, Z., 2009. Quantile cointegrating regression. J. Econometrics 150 (2), 248–260.
Yang, B., Liu, X., Peng, L., Cai, Z., 2021. Unified tests for a dynamic predictive regression. J. Bus. Econom. Statist. 39 (3), 684–699.
Yang, B., Long, W., Peng, L., Cai, Z., 2019. Testing the predictability of US housing price index returns based on an IVX-AR model. J. Amer. Statist. Assoc. 115 (532), 1598–1619.
Zhu, F., Cai, Z., Peng, L., 2014. Predictive regressions for macroeconomic data. Ann. Appl. Stat. 8 (1), 577–594.