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] MOE Key Laboratory of Econometrics, and Wang Yanan Institute for Studies in Economics, 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.

April 30, 2020

Abstract: For predictive quantile regressions with highly persistent regressors, a conventional test statistic suffers from a serious size distortion and its limiting distribution relies on the unknown persistence degree of predictors. This paper proposes a double-weighted approach to offer a robust inferential theory across all types of persistent regressors. 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 nice 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 persistence degrees in multiple regressions. Simulations and empirical studies are provided to demonstrate the effectiveness of the newly proposed approach. The heterogenous predictability of US stock returns at different quantile levels is reexamined.

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

---

*The authors would like to thank Professor Ji-Hyung Lee for sharing the empirical dataset and code for IVX-QR. This work, in part, was supported by the National Natural Science Foundation of China with grant numbers #71631004 (Key Project), #71571152 and #71850011, and the Fundamental Research Funds for the Central Universities with grant numbers #20720181004, #20720171002 and #JBK2001034, and funds provided by Fujian Provincial Key Laboratory of Statistics (Xiamen University) #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.

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

1 Introduction

A long-term issue in financial statistics is to test whether or not a return process (say, asset return or housing price return) is predictable by a set of lagged predictors (say, financial ratios or/and macroeconomic variables). The typical method in previous studies is an ordinary least squares (OLS) approach, which is applied to mean regressions, while conventional test statistics are used to test the significance of coefficients. The conclusions 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), Campbell and Thompson (2008), Welch and Goyal (2008), Rapach, Strauss and Zhou (2010), Sekkel (2011), and the references therein. The indefinite conclusions are partially due to the statistical issues caused by those highly persistent regressors where conventional test statistics are invalid with a serious size distortion. The problem is more serious if the innovation in the predictor is highly correlated with the innovation in dependent variable, which is the so-called *embedded endogeneity*, as studied by Campbell and Yogo (2006), Torous, Valkanov and Yan (2004), Zhu, Cai and Peng (2014), Yang, Long, Peng and Cai (2019), and among others.[1] Another explanation is that the predictability of asset returns might be heterogenous, relying on economic environment. For example, a stronger prediction power is usually found in recession periods for stock markets; see Gonzalo and Pitarakis (2012), which implies potentially greater predictability at lower quantiles. Because 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) and Xiao (2009) for details.

Testing the predictability in a quantile setting is of importance in economics and statistics 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

[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), Kostakis, Magdalinos and Stamatogiannis (2015), Phillips and Lee (2016), and Yang et al. (2019), the weighted empirical likelihood approach by Zhu, Cai and Peng (2014), Liu, Yang, Cai and Peng (2019), and Yang, Liu, Peng and Cai (2018), and the variable addition (VA) or augmented regression or control function approach by Elliott (2011) and Breitung and Demetrescu (2015) and Yang et al.(2018).

future stock returns is informative for investment decisions of risk averse investors. Second, from the statistical point of view, quantile regressions are more suitable when the distribution is skewed and/or heavy tailed, which is a stylized fact in financial statistics, and consequently 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 is not a trivial task. The main challenging statistical issues in mean regressions causing the failure of traditionally statistical inferences of the predictive regression still exist for predictive quantile regressions. 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 predictive quantile regressions with both various degrees of persistency and embedded endogeneity. Indeed, Lee (2016) extended the exogenous instrumental variable approach filtering methodology by Magdalinos and Phillips (2009), Kostakis et al.(2015) for mean regressions to quantile regression, 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, which can be applied to the multiple predictors case. Recently, Fan and Lee (2019) extended the IVX-QR method in Lee (2016) to the situation with conditionally heteroskedastic errors. However, the IVX-QR requires that the instrumental variable should be less persistent than the predictors. Thus, it might lose some of its test power as illustrated in Kostakis et al. (2015). Meanwhile, the performance of the test is sensitive to the choice of turning parameters involved in the construction of mildly integrated instrumental variables, and it is difficult to extend to the case with mixed persistent regressors, including both stationary and nonstationary predictors.

The main contribution of this paper is to propose a novel approach, termed as double weighted method, to develop a uniform inferential theory for predictive quantile regressions with highly persistent variables. Our 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 is well-selected through a data-driven approach, such that the auxiliary regressor enjoys having the same persistency degree with the original predictor. Using the coefficients of both original regressor and auxiliary regressor, with a similar idea of rotation, we construct a weighted estimator between them to eliminate the impact of the embedded endogeneity. 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 does not require a less persistent instrumental variable, and it could reach the local power under the optimal convergence rate T with nonstationary predictors and $\sqrt{T}$ with stationary predictors, respectively. More importantly, our method can easily be generalized to multiple regressors with mixed persistence degrees and this generalization is seminal in the 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 over IVX-QR method.

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 indictors. For the convenience of comparison, the same data set used by Lee (2016) is taken with the sample period from 1927 to 2005. To view 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 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 result of this empirical study is reported in Section 6.

Our paper is closely related to the literature of predictive regression with highly persistent regressors. Acknowledging the fact that the asymptotic distribution relies on the time series properties of the regressors and errors, a series of research papers have aimed to developing a uniform inference theory on predictive mean regressions in the sense that the testing procedure for testing predictability is robust to different persistence categories, including, but not limited to, the papers by Campbell and Yogo (2006), Magdalinos and Phillips (2009), Chen and Deo (2009), Chen, Deo and Yi (2013), Phillips and Lee (2013), Zhu et al. (2014), Kostakis et al. (2015), Phillips and Lee (2016), Yang et al. (2018), Yang et al. (2019), and Liu et al. (2019), which focused on predictive mean regression models.

Also, 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 variable addition approach in the context of vector autoregressive processes with unknown persistence. In particular, Breitung and Demetrescu (2015) argued that the traditional VA approaches suffer from a loss of power and generalized VA approach by using instrumental variables that are constructed exogenously or endogenously. 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 Section 3 provides the procedures to estimate parameters and to construct the test statistics and also presents the asymptotic theories for the proposed estimators and the test statistics. An extension to the multiple regressors with mixed persistence degrees is discussed in Section 4. Section 5 reports the Monte Carlo simulation results. Section 6 presents the analysis results for the empirical applications. Finally, Section 7 concludes the paper. The detailed proofs of the main results are given in Appendix.

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 Model Framework

Assume that y_t is a dependent variable and its τth quantile is $Q_{y_t}(\tau|\mathcal{F}_{t-1})$, defined by $P(y_t \leq Q_{y_t}(\tau|\mathcal{F}_{t-1})|\,\mathcal{F}_{t-1}) = \tau$, 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

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = Q_{y_t}(\tau|x_{t-1}) = \mu_\tau + \beta_\tau x_{t-1}, \tag{2.1}$$

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 or macroeconomic variable and so on, which is a time series, commonly modeled by an autoregressive (AR) model as

$$x_t = \rho x_{t-1} + v_t,\ \rho = 1 + c/T^\alpha,\ \ 1 \leq t \leq T, \tag{2.2}$$

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 (2.2). For simplicity of exposition, 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 in the literature:

(I0) stationary: $\alpha = 0$ and $|1 + c| < 1$;

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

(NI1) local to unit root: $\alpha = 1$ and $c < 0$;

(I1) unit root: $c = 0$;

(LE) local to unity on the explosive side: $\alpha = 1$ and $c > 0$.

Of course, it is interesting to consider the other cases as $0 < \alpha < 1$, corresponding to the so-called mildly integrated processes ($c < 0$) or mildly explosive processes ($c > 0$). The latter can be used to explore the mild economic or financial bubbles and other applications, see Phillips, Shi and Yu (2015) and the references therein.[3] Here, following Lee (2016), a general weakly dependent innovation structure of the linear process on $\{v_t\}$ in (2.2) is imposed and listed below.

Assumption 2.1. *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 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 2.1. Assumptions 2.1 allows for linear process dependence for v_t and imposes a conditionally homoskedastic martingale difference sequence (mds) condition for ε_t. Different from Lee (2016), here we do not specify a linear predictive mean regression model and hence avoid to impose any assumption on the innovation for the mean regression model. Note that, for the univariate case, $K = 1$.

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 also, $\psi_\tau(u_{t\tau}) = \tau - 1(u_{t\tau} < 0)$. Now, 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 $E[\psi_\tau(u_{t\tau})^4] = -3\tau^4 + 6\tau^3 - 4\tau^2 + \tau < \infty$. One may refer to Appendix for the details of proof. Further, define

$$\Sigma_{\psi_\tau v} = \sum_{h=-\infty}^{\infty} E[\psi_\tau(u_{t\tau})v_{t+h}] = F_x(1)E[\psi_\tau(u_{t\tau})\varepsilon_t].$$

By Lemma A.2 in Appendix, 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

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\begin{pmatrix}\psi_\tau(u_{t\tau})\\ v_t\end{pmatrix} \Rightarrow \begin{pmatrix}B_{\psi_\tau}(r)\\ B_v(r)\end{pmatrix} = BM\begin{pmatrix}\tau(1-\tau) & \Sigma_{\psi_\tau v}\\ \Sigma_{\psi_\tau v} & \Omega_{vv}\end{pmatrix}, \tag{2.3}$$

[3] Our methods can be extended to allow for these two cases with some adjustment. To make the proof easy to follow, our focus is on the simple setting.

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

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

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

where Z represents the standard normal distributions. 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; see also Fan and Lee (2019) for more details, because it is extremely challenging to estimate consistently the nuisance parameter c and to test if the persistence α equals zero or not or $0 < \alpha < 1$. Thus, it is necessary to develop a unified inference method to avoid the mistake of making a false judgement about the persistence of predictors under a quantile framework.

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

Assumption 2.2. *(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 $E(f^{\vartheta}_{u_{t\tau},t-1}(0)) < \infty$ 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 2.2. As shown by Xiao (2009), the above conditions in Assumption 2.2 are quite standard and not restrictive. In particular, the part (i) in Assumption 2.2 is not as restrictive as the counterpart assumption in Lee (2016), which assumes that $f_{u_{t\tau},t-1}(0)$ follows the FCLT.

3 Statistical Modeling Procedures

3.1 Estimation Approach

Motivated by the variable addition 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. That is, (2.1) is re-written as follows:

$$Q_{y_t}(\tau|x_{t-1}) = \mu_\tau + \beta_\tau x_{t-1} = \mu_\tau + \beta_\tau x^*_{t-1} + \gamma_\tau z_{t-1}, \tag{3.1}$$

where $x^*_{t-1} = x_{t-1} - z_{t-1}$ and z_{t-1} is an additional (auxiliary) variable which is chosen in Section 3.2 in detail. Note that $\gamma_\tau = \beta_\tau$ in (3.1), which will be used to construct weighted combined estimator for β_τ later. Clearly, μ_τ, β_τ and γ_τ in (3.1) can be estimated by running the following quantile regression

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

where $\rho_\tau(u) = u[\tau - 1(u < 0)]$ is the so-called check function in the statistics literature. Note that Breitung and Demetrescu (2015) only used $\hat{\gamma}_\tau$, the estimator of the coefficient of the auxiliary variable z_t, to construct the test statistic in the predictive mean regression, and required z_t to be an instrumental variable (IV) less persistent than x_t or an exogenous deterministic or stochastic trend process, in order to guarantee that the asymptotic distribution of the test statistic is irrelevant to the nuisance parameter c. However, if z_t is generated as an IV less persistent than x_t, the corresponding test statistic suffers from the loss of power for the case with nonstationary x_t, while if z_t is generated as an exogenous deterministic or stochastic trend process, the test is invalid for the case with stationary x_t.

To avoid this problem, the variable addition 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 persistence 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 constructed by self-normalization to eliminate the nuisance parameter c, can avoid not only the size distortion but also the loss of power with arbitrary persistence.

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 will be presented in Section 3.2. As mentioned earlier, $\gamma_\tau = \beta_\tau$ so 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 here to construct the estimator for β_τ, which is the weighted sum of $\hat{\beta}_\tau$ and $\hat{\gamma}_\tau$, denoted by $\hat{\beta}^w_\tau$,

$$\hat{\beta}^w_\tau = \frac{W_1}{W_1 + W_2}\hat{\beta}_\tau + \frac{W_2}{W_1 + W_2}\hat{\gamma}_\tau, \tag{3.2}$$

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[4] and is irrelevant to

[4]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.

the nuisance parameter c after normalization. For this purpose, the weights W_1 and W_2 are taken to be

$$W_1 = \sum_{t=2}^{T} x^*_{t-1} z_{t-1}/T^2 - \sum_{t=2}^{T} x^*_{t-1} \sum_{t=2}^{T} z_{t-1}/T^3, \tag{3.3}$$

and

$$W_2 = \sum_{t=2}^{T} z^2_{t-1}/T^2 - \left(\sum_{t=2}^{T} z_{t-1}\right)^2 /T^3. \tag{3.4}$$

Note that in Section 3.3, some arguments will be provided to explain the reason on why the above W_1 and W_2 are used.

3.2 Choice of Auxiliary Variable

This section is devoted to how to construct 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 $W_{\zeta,T}(r) = \zeta_{[rT]}/\sqrt{T}$ for $0 \le r \le 1$ and $B(\cdot)$ is the standard Brownian motion. In the second step, the coefficient $\hat{\pi}_1$ is obtained by estimating the following regression

$$x_{t-1} = \pi_0 + \pi_1 \zeta_{t-1} + e_t. \tag{3.5}$$

Finally, we define z_{t-1} as a linear combination of ζ_{t-1} and one bounded transformation of x_{t-1} as follows

$$z_{t-1} = \hat{\pi}_1 \zeta_{t-1} + x_{t-1}/\sqrt{1 + x^2_{t-1}}. \tag{3.6}$$

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 3.1. 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 Phillips and Magdalinos (2009). Type II instruments include strictly exogenous nonstationary variables, deterministic terms and Cauchy type instrument. Therefore, in a certain 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}. However, the random walk instrument ζ_{t-1} does not work for stationary cases, while $x_{t-1}/\sqrt{1 + x^2_{t-1}}$ can not handle the predictive regression with intercept term for nonstationary cases without some necessary adjustments.[5] 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 (3.5). By doing so, our method is robust to both nonstationary and stationary cases, and can be easily extended to the multivariate case with mixed persistence.

The following proposition can be established for the asymptotic properties of $\hat{\pi}_1$.

Proposition 3.1. It follows that

$$\hat{\pi}_1 = \sum_{t=2}^{T} \bar{\zeta}_{t-1}\bar{x}_{t-1} / \sum_{t=2}^{T} \bar{\zeta}_{t-1}^2 = \begin{cases} \tilde{\pi}_1 + o_p(1), & \text{NI1, I1 and LE;} \\ O_p(T^{-1}), & \text{I0,} \end{cases} \tag{3.7}$$

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 $\tilde{\pi}_1 = \int \bar{B}(r)\bar{J}_x^c(r)dr / \int \bar{B}(r)^2 dr$ with $\bar{B}(r) = B(r) - \int B(r)dr$ and $\bar{J}_x^c(r) = J_x^c(r) - \int J_x^c(r)dr$.

Remark 3.2. The proof is standard and thus, details are skipped here. Clearly, (3.7) implies that 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. Particularly, if x_{t-1} is nonstationary, including NI1, I1 and LE, $\hat{\pi}_1$ converges to a nonzero random variable due to the spurious correlation between x_{t-1} and ζ_{t-1} (Phillips, 2014), and the second term $x_{t-1}/\sqrt{1+x_{t-1}^2}$ is dominated by the first term $\hat{\pi}_1\zeta_{t-1}$. If x_{t-1} is stationary, then $\hat{\pi}_1$ converges to zero with the convergence rate T and the first term in z_{t-1} is dominated by the second term $x_{t-1}/\sqrt{1+x_{t-1}^2}$.

Moreover, given the above construction of z_{t-1}, the asymptotic property of W_1+W_2 can be established easily for the cases with stationary and nonstationary x_t, respectively.

Proposition 3.2. It is easy to show that

$$\begin{cases} T(W_1+W_2) = E\left[x_t^2(1+x_t^2)^{-1/2}\right] + o_p(1), & \text{I0;} \\ W_1+W_2 = \pi_c^2 + o_p(1), & \text{NI1, I1 and LE,} \end{cases} \tag{3.8}$$

where $\pi_c = \int \bar{B}(r)\bar{J}_x^c(r)dr \left[\int \bar{B}^2(r)dr\right]^{-1/2}$ with $\tilde{\pi}_1$ defined in Proposition 3.1.

Remark 3.3. The basic idea for showing the above proposition is as follows. If x_{t-1} is nonstationary, by plugging (3.6) into (3.3) and (3.4), one can show easily that

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

where $\bar{z}_{t-1} = z_{t-1} - \sum_{t=2}^{T} z_{t-1}/T$. On the other hand, if x_{t-1} is stationary, z_{t-1} is determined by $x_{t-1}/\sqrt{1+x_{t-1}^2}$ and then,

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

[5]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 and loses the power of test.

3.3 Large Sample Theory

To obtain the asymptotic distribution of $\hat{\beta}^w_\tau$, we will first establish the so-called Bahadur representation[6] for $\hat{\boldsymbol{\theta}}_\tau$; that is, use the first order approximation to get an explicit expression for $\hat{\boldsymbol{\theta}}_\tau$. To this end, define $\hat{\boldsymbol{\theta}}^a_\tau = \boldsymbol{D}_T(\hat{\mu}_\tau - \mu_\tau, \hat{\beta}_\tau - \beta_\tau, \hat{\gamma}_\tau - \beta_\tau)^\top$, where $\boldsymbol{D}_T = \mathrm{diag}(\sqrt{T}, T, T)$ for NI1, I1 and LE and $\boldsymbol{D}_T = \mathrm{diag}(\sqrt{T}, \sqrt{T}, \sqrt{T})$ for I0. Then, the Bahadur representation for $\hat{\boldsymbol{\theta}}^a_\tau$ is given as follows with its mathematical proof given in Appendix. Note that this result is new in the literature when regressors might be nonstationary and is of own interest.

Theorem 3.1. *(Bahadur Representation) Under Assumptions 2.1 and 2.2,*

$$\hat{\boldsymbol{\theta}}^a_\tau = f_{u_\tau}(0)^{-1} N_T^{-1} \boldsymbol{D}_T^{-1} \sum_{t=2}^{T} \Lambda_{t-1} \psi_\tau(u_{t\tau}) + o_p(1), \tag{3.9}$$

where $\Lambda_{t-1} = (1, x^*_{t-1}, z_{t-1})^\top$, $N_T = \boldsymbol{D}_T^{-1} \sum_{t=2}^T \Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1}$, *and* $f_{u_\tau}(0)$ *is defined in Assumption 2.2 (i).*

Remark 3.4. From Theorem 3.1, one can see clearly that the second and the third components of the vector on the right-hand side of (3.9) involves x^*_{t-1}. To construct a pivotal test statistic free of nuisance parameter c, the weighted estimator $\hat{\beta}^w_\tau$ is constructed as in (3.2), with a similar idea of rotation in factor analysis, to get rid of x^*_{t-1}. It will then be shown by Lemma A.5 in Appendix that the following result holds true for $\hat{\beta}^w_\tau$,

$$(W_1 + W_2) T(\hat{\beta}^w_\tau - \beta_\tau) = f_{u_\tau}(0)^{-1} \sum_{t=2}^{T} \frac{1}{\sqrt{T}} \left(z_{t-1} - \sum_{t=2}^{T} z_{t-1}/T \right) \psi_\tau(u_{t\tau})/\sqrt{T} + o_p(1). \tag{3.10}$$

Evidently, in contrast from the second or the third components of the vector on the right-hand side of (3.9), the right-hand side of (3.10) involves only z_{t-1} but not x_{t-1} or x^*_{t-1} so that it makes the asymptotic (or mixture) normality of $\hat{\beta}^w_\tau$ only depends on z_{t-1}.

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

Theorem 3.2. *Under Assumptions 2.1 and 2.2, for I0, NI1, I1 and LE cases, the asymptotic distribution of* $\hat{\beta}^w_\tau$ *is given below,*

$$\begin{cases} \sqrt{T}(\hat{\beta}^w_\tau - \beta_\tau) \xrightarrow{d} \mathrm{N}\left(0, \sigma^2_{\beta_\tau}\right), & \textit{I0}, \\ T\,\pi_c\,(\hat{\beta}^w_\tau - \beta_\tau) \xrightarrow{d} \mathrm{N}\left(0, \sigma^2_\tau\right), & \textit{NI1, I1 and LE}, \end{cases} \tag{3.11}$$

where with $\sigma^2_\tau = \tau(1-\tau)/f^2_{u_\tau}(0)$, $\sigma^2_{\beta_\tau} = \sigma^2_\tau \left\{E\left[x_t^2(1+x_t^2)^{-1/2}\right]\right\}^{-2} Var\left[x_t(1+x_t^2)^{-1/2}\right]$ *and* π_c *is given in Proposition 3.2*

[6] See, for example, Cai and Xu (2008) for stationary quantile regression.

Remark 3.5. Clearly, Theorem 3.2 shows the convergence rate of the estimator of $\hat{\beta}^w_\tau$ with N1, I1 and LE x_t is faster than that for the IVX-QR method proposed in Lee (2016).

Although the asymptotic distribution of $\hat{\beta}^w_\tau$ with NI1, I1 and LE x_t still contains the nuisance parameter c, we can construct the t-test statistic t^w by self normalization because the asymptotic distribution of $\hat{\beta}^w_\tau$ is mixture normal, as follows:

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

where $\hat{f}_{u_\tau}(0)$ is a consistent estimator of $f_{u_\tau}(0)$, while 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.

Theorem 3.3. *(1) Under the null hypothesis $H_0 : \beta_\tau = 0$,*

$$t^w \xrightarrow{d} \mathrm{N}(0,1).$$

(2)(a) Under the local alternative hypothesis $H_a : \beta_\tau = b_\tau/\sqrt{T}$ for any b_τ, if x_{t-1} is I0,

$$t^w \xrightarrow{d} N(b_\tau/\sigma_{\beta_\tau}, 1),$$

where σ_{β_τ} is defined in Theorem 3.2.
(b) Under the local alternative hypothesis $H_a : \beta_\tau = b_\tau/T$ for any b_τ, if x_{t-1} is NI1, I1 or LE,

$$t^w \xrightarrow{d} N(0,1) + b_\tau|\pi_c|/\sigma_\tau,$$

where π_c is given in Proposition 3.2 and σ_τ is defined in Theorem 3.2.

Remark 3.6. From Theorem 3.3, 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).

4 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; 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, at al. (2014), Liu et al. (2019), and Yang et al. (2019) can not be easily extended to a multiple regression. However, the proposed method in previous section can be easily extended to the following multivariate predictive quantile regression with mixed persistencies

$$Q_{y_t}(\tau|\boldsymbol{X}_{t-1}) = \mu_\tau + \boldsymbol{\beta}_\tau^\top \boldsymbol{X}_{t-1}, \tag{4.1}$$

where $\boldsymbol{\beta}_\tau = (\beta_{1\tau}, \beta_{2\tau}, \cdots, \beta_{K\tau})^\top$ is a $K \times 1$ vector and $\boldsymbol{X}_{t-1}$ is a $K \times 1$ vector of predictors, which might contain both stationary and nonstationary predictors. For the purpose of exposition, $\boldsymbol{X}_{t-1}$ is written as $\boldsymbol{X}_{t-1} = (\boldsymbol{X}_{1,t-1}^\top, \boldsymbol{X}_{2,t-1}^\top)^\top$ with $\boldsymbol{X}_{1,t-1} = (x_{1,t-1}, x_{2,t-1} \cdots, x_{K_1,t-1})^\top$ being nonstationary and $\boldsymbol{X}_{2,t-1} = (x_{K_1+1,t-1}, x_{K_1+2,t-1}, \cdots, x_{K,t-1})^\top$ being stationary. It is assumed there is no cointegration relationship among $\boldsymbol{X}_{1,t-1}$. Note that $0 \le K_1 \le K$ and $K_1 = 0$ means all elements in $\boldsymbol{X}_{t-1}$ are I0, while $K_1 = K$ means all elements in $\boldsymbol{X}_{t-1}$ are NI1, I1 or LE. Now, $x_{i,t}$ can be modeled by an AR(1) as

$$x_{i,t} = \rho_i x_{i,t-1} + v_{i,t}, \quad \rho_i = \begin{cases} 1 + c_i/T, & i = 1, \cdots, K_1; \\ 1 + c_i, \text{ where } |1 + c_i| < 1, & i = K_1 + 1, \cdots, K \end{cases} \tag{4.2}$$

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 $i = 1, \cdots, K_1$, $x_{i,\lfloor rT \rfloor}/\sqrt{T} \Rightarrow J_{x_i}^{c_i}(r)$ 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 i-th element of $B_v(s)$, which is a vector of Brownian motions defined in (2.3).

Remark 4.1. It is clear that the model in (4.1) is new and it covers some known models in mean models in the literature. For example, if there is nonstationary part ($K_1 = 0$), (4.1) reduces to the model studied by Amihud, Hurvich and Wang (2009) for mean regression models.

To estimate μ_τ and $\boldsymbol{\beta}_\tau$ in (4.1), let $\boldsymbol{X}_{t-1}^* = \boldsymbol{X}_{t-1} - \boldsymbol{Z}_{t-1}$ and $\boldsymbol{Z}_{t-1}$ be the vector of additional variables. Then, μ_τ and $\boldsymbol{\beta}_\tau$ can be estimated based on the variable addition as follows:

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

where $\boldsymbol{Z}_t = (z_{1,t}, z_{2,t} \cdots, z_{K,t})^\top$ is constructed by three steps similar to the univariate case as in Section 3.2; 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 $\boldsymbol{X}_t$. 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, we run the following quantile regression:

$$x_{i,t} = \pi_{0,i} + \pi_{1,i}\, \zeta_{i,t-1} + e_{i,t}$$

for all $1 \le i \le K$. Similarly, one can show that $\hat{\pi}_{1,i} \xrightarrow{d} \tilde{\pi}_{1,i} = \int \bar{B}_i(r)\bar{J}_{x_i}^{c_i}(r)dr / \int \bar{B}_i(r)^2 dr$, 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 follows

$$z_{i,t-1} = \hat{\pi}_{1,i}\zeta_{i,t-1} + x_{i,t-1}/\sqrt{1 + x_{i,t-1}^2}.$$

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 persistence.

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

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

where

$$\boldsymbol{W}_1 = \sum_{t=2}^{T} \boldsymbol{Z}_{t-1}(\boldsymbol{X}_{t-1}^*)^\top/T^2 - \sum_{t=2}^{T} \boldsymbol{Z}_{t-1}\sum_{t=2}^{T}(\boldsymbol{X}_{t-1}^*)^\top/T^3,$$

and

$$\boldsymbol{W}_2 = \sum_{t=2}^{T} \boldsymbol{Z}_{t-1}\boldsymbol{Z}_{t-1}^\top/T^2 - \sum_{t=2}^{T} \boldsymbol{Z}_{t-1}\sum_{t=2}^{T}\boldsymbol{Z}_{t-1}^\top/T^3.$$

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 persistence cases, we define the following weighting matrix $\boldsymbol{D}_T$ accordingly

$$\boldsymbol{D}_T = \begin{cases} \mathrm{diag}(\sqrt{T}, \sqrt{T}), & K_1 = 0; \\ \mathrm{diag}(T, \sqrt{T}), & K_1 = 1; \\ \mathrm{diag}(T, T), & K_1 = 2. \end{cases}$$

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

Case 1 ($K_1 = 0$):

$$\boldsymbol{V}_1 = \begin{pmatrix} E\left(\frac{x_{1,t}^2}{\sqrt{1+x_{1,t}^2}}\right) & E\left(\frac{x_{1,t}x_{2,t}}{\sqrt{1+x_{1,t}^2}}\right) \\ E\left(\frac{x_{1,t}x_{2,t}}{\sqrt{1+x_{2,t}^2}}\right) & E\left(\frac{x_{2,t}^2}{\sqrt{1+x_{2,t}^2}}\right) \end{pmatrix}, \tag{4.3}$$

and

$$\boldsymbol{V}_2 = \begin{pmatrix} E\left(\frac{x_{1,t}^2}{1+x_{1,t}^2}\right) - E\left(\frac{x_{1,t}}{\sqrt{1+x_{1,t}^2}}\right)^2 & E\left(\frac{x_{1,t}}{\sqrt{1+x_{1,t}^2}}\frac{x_{2,t}}{\sqrt{1+x_{2t}^2}}\right) - E\left(\frac{x_{1,t}}{\sqrt{1+x_{1,t}^2}}\right)E\left(\frac{x_{2,t}}{\sqrt{1+x_{2,t}^2}}\right) \\ E\left(\frac{x_{1,t}}{\sqrt{1+x_{1,t}^2}}\frac{x_{2,t}}{\sqrt{1+x_{2t}^2}}\right) - E\left(\frac{x_{1,t}}{\sqrt{1+x_{1,t}^2}}\right)E\left(\frac{x_{2,t}}{\sqrt{1+x_{2,t}^2}}\right) & E\left(\frac{x_{2,t}^2}{1+x_{2,t}^2}\right) - E\left(\frac{x_{2,t}}{\sqrt{1+x_{2,t}^2}}\right)^2 \end{pmatrix}. \tag{4.4}$$

Case 2 ($K_1 = 1$):

$$\boldsymbol{V}_1 = \begin{pmatrix} \tilde{\pi}_{1,1} \int \bar{B}_1(r) J_{x_1}^{c_1}(r) dr & 0 \\ 0 & E\left(\frac{x_{2,t}^2}{\sqrt{1+x_{2,t}^2}}\right) \end{pmatrix}, \tag{4.5}$$

and

$$\boldsymbol{V}_2 = \begin{pmatrix} \tilde{\pi}_{1,1}^2 \int \bar{B}_1(r)^2 dr & 0 \\ 0 & E\left(\frac{x_{2,t}^2}{1+x_{2,t}^2}\right) - E\left(\frac{x_{2,t}}{\sqrt{1+x_{2,t}^2}}\right)^2 \end{pmatrix}. \tag{4.6}$$

Case 3 ($K_1 = 2$):

$$\boldsymbol{V}_1 = \begin{pmatrix} \tilde{\pi}_{1,1} \int \bar{B}_1(r) J_{x_1}^{c_1}(r) dr & \tilde{\pi}_{1,1} \int \bar{B}_1(r) J_{x_2}^{c_2}(r) dr \\ \tilde{\pi}_{1,2} \int \bar{B}_2(r) J_{x_1}^{c_1}(r) dr & \tilde{\pi}_{1,2} \int \bar{B}_2(r) J_{x_2}^{c_2}(r) dr \end{pmatrix}, \tag{4.7}$$

and

$$\boldsymbol{V}_2 = \begin{pmatrix} \tilde{\pi}_{1,1}^2 \int \bar{B}_1(r)^2 dr & \tilde{\pi}_{1,1}\tilde{\pi}_{1,2} \int \bar{B}_1(r)\bar{B}_2(r) dr \\ \tilde{\pi}_{1,1}\tilde{\pi}_{1,2} \int \bar{B}_2(r)\bar{B}_1(r) dr & \tilde{\pi}_{1,2}^2 \int \bar{B}_2(r)^2 dr \end{pmatrix}. \tag{4.8}$$

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

Theorem 4.1. *Under Assumptions 2.1 and 2.2, the asymptotic distribution of $\hat{\boldsymbol{\beta}}_\tau^w$ is given by*

$$\begin{aligned} &\boldsymbol{D}_T(\hat{\boldsymbol{\beta}}_\tau^w - \boldsymbol{\beta}_\tau) \\ &= f_{u_\tau}(0)^{-1} \left[(\boldsymbol{D}_T)^{-1} \sum_{t=2}^{T} \left(\boldsymbol{Z}_{t-1} - \frac{1}{T} \sum_{t=2}^{T} \boldsymbol{Z}_{t-1} \right) \boldsymbol{X}_{t-1}^\top (\boldsymbol{D}_T)^{-1} \right]^{-1} (\boldsymbol{D}_T)^{-1} \sum_{t=2}^{T} \left(\boldsymbol{Z}_{t-1} - \frac{1}{T} \sum_{t=2}^{T} \boldsymbol{Z}_{t-1} \right) \psi_\tau(u_{t\tau}) + o_p(1) \\ &\xrightarrow{d} f_{u_\tau}(0)^{-1} \boldsymbol{V}_1^{-1} \, \mathrm{MN}\left(0, \tau(1-\tau)\boldsymbol{V}_2\right), \end{aligned} \tag{4.9}$$

where $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are defined in (4.3)-(4.8), respectively.

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

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

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 proof given in Appendix.

Theorem 4.2. *Under Assumptions 2.1 and 2.2 and the null hypothesis H_0: $R\boldsymbol{\beta}_\tau = \boldsymbol{r}_\tau$, one has*

$$Q_m^w \xrightarrow{d} \chi_r^2,$$

where χ_r^2 is a χ^2-distribution with r degrees of freedom.

5 Monte Carlo Simulation Studies

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 persistence which is not studied by Lee (2016). For the first simulations, a comparison with the IVX-QR approach in Lee (2016) is reported.

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

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

where $\rho = 1 + c/T^{\alpha}$. To create the embedded endogeneity among innovations, the innovation processes are generated as $(u_t, v_t)^{\top} \sim iid\, N(0_{2\times 1}, \Sigma_{2\times 2})$, where $\Sigma = \begin{pmatrix} 1 & -0.95 \\ -0.95 & 1 \end{pmatrix}$. By Proposition 1 of Gaglianone, Lima, Linton and Smith (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 τ-th quantile of u_t.

First, the results for the comparison of the size performances of our method with IVX-QR for two cases, $\alpha = 1$ and $\alpha = 0$, are shown in Tables 1 for $\alpha = 1$ and 2 for $\alpha = 0$, respectively, with sample sizes of $T = 150$, 300 and 700, the different values of τ as $\tau = 0.05$, 0.1, $\cdots$, 0.90, and 0.95, and the nominal size at 5%. Simulation is repeated 100 times for each setting and the rejection rate is computed based on 500 simulations. The mean and the standard error in parenthesis of 100 rejection rates are given in Tables 1 for $\alpha = 1$ and 2 for $\alpha = 0$, respectively. For each setting, four values of c are considered further as $c = 1.5$, 0, -5 and -25, corresponding to the cases: LE, I1, NI1 and NI1 (with large deviation from unit root). Clearly, the following findings can be evidently observed from Table 1 for $\alpha = 1$. First, 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 still suffers from somehow size distortion, where there is a over-rejection for the case of LE predictors, and an under-rejection for the some case of NI1 predictors ($c = -5$). Second, it is not surprising to see that due to less data points in tails, both methods might have size distortions for the extreme quantile levels, but for most cases, the newly proposed method performs better. Similar findings can be summarized for the stationary case, i.e., $\alpha = 0$ from Table 2, and for most cases, the IVX-QR has an under-rejection problem, because it invalids the requirement for a less persistent instrumental variable.

Table 1: Size performances of t^w and IVX-QR for $\alpha = 1$ with the nominal size 5%.

						t^w						
τ		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
T=150	c=1.5	0.114	0.081	0.066	0.060	0.059	0.059	0.057	0.061	0.065	0.083	0.108
		(0.014)	(0.013)	(0.011)	(0.012)	(0.011)	(0.011)	(0.010)	(0.010)	(0.011)	(0.013)	(0.013)
	c=0	0.110	0.082	0.063	0.057	0.055	0.054	0.056	0.059	0.062	0.080	0.111
		(0.014)	(0.011)	(0.012)	(0.010)	(0.012)	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)	(0.014)
	c=-5	0.108	0.075	0.057	0.051	0.050	0.050	0.050	0.052	0.059	0.078	0.107
		(0.013)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.010)	(0.012)	(0.013)
	c=-25	0.109	0.077	0.058	0.054	0.051	0.049	0.051	0.055	0.059	0.078	0.111
		(0.013)	(0.012)	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.011)	(0.012)	(0.011)	(0.015)
T=300	c=1.5	0.086	0.071	0.060	0.056	0.055	0.055	0.056	0.057	0.062	0.067	0.087
		(0.013)	(0.012)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)	(0.012)
	c=0	0.085	0.069	0.057	0.055	0.053	0.053	0.052	0.053	0.058	0.067	0.084
		(0.013)	(0.011)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.011)	(0.011)	(0.012)
	c=-5	0.084	0.065	0.053	0.048	0.048	0.048	0.048	0.050	0.052	0.063	0.083
		(0.011)	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.011)	(0.012)
	c=-25	0.089	0.069	0.056	0.052	0.051	0.051	0.050	0.054	0.056	0.069	0.089
		(0.015)	(0.011)	(0.009)	(0.010)	(0.011)	(0.009)	(0.010)	(0.011)	(0.009)	(0.010)	(0.013)
T=700	c=1.5	0.070	0.062	0.055	0.053	0.052	0.052	0.052	0.053	0.056	0.062	0.073
		(0.011)	(0.011)	(0.009)	(0.010)	(0.010)	(0.010)	(0.008)	(0.011)	(0.011)	(0.013)	(0.012)
	c=0	0.070	0.059	0.053	0.051	0.049	0.049	0.050	0.051	0.050	0.062	0.071
		(0.013)	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)	(0.011)	(0.010)	(0.009)	(0.010)	(0.010)
	c=-5	0.069	0.057	0.049	0.049	0.046	0.048	0.047	0.049	0.052	0.058	0.069
		(0.012)	(0.011)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
	c=-25	0.071	0.060	0.055	0.054	0.052	0.049	0.052	0.052	0.056	0.062	0.073
		(0.011)	(0.009)	(0.011)	(0.009)	(0.011)	(0.010)	(0.010)	(0.011)	(0.009)	(0.011)	(0.012)
						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
T=150	c=1.5	0.175	0.130	0.113	0.155	0.102	0.090	0.101	0.152	0.114	0.129	0.179
		(0.017)	(0.016)	(0.014)	(0.017)	(0.014)	(0.014)	(0.015)	(0.017)	(0.013)	(0.014)	(0.016)
	c=0	0.138	0.096	0.074	0.109	0.071	0.059	0.069	0.105	0.074	0.093	0.136
		(0.015)	(0.013)	(0.012)	(0.017)	(0.012)	(0.011)	(0.013)	(0.015)	(0.012)	(0.012)	(0.015)
	c=-5	0.098	0.063	0.043	0.047	0.038	0.035	0.038	0.047	0.044	0.061	0.095
		(0.012)	(0.012)	(0.011)	(0.010)	(0.009)	(0.008)	(0.008)	(0.010)	(0.010)	(0.010)	(0.013)
	c=-25	0.115	0.085	0.063	0.058	0.054	0.051	0.052	0.056	0.064	0.085	0.114
		(0.015)	(0.010)	(0.011)	(0.010)	(0.011)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)	(0.013)
T=300	c=1.5	0.162	0.135	0.119	0.164	0.097	0.093	0.096	0.162	0.118	0.134	0.166
		(0.017)	(0.015)	(0.013)	(0.016)	(0.014)	(0.013)	(0.012)	(0.016)	(0.015)	(0.015)	(0.017)
	c=0	0.119	0.093	0.074	0.108	0.056	0.054	0.058	0.110	0.072	0.088	0.119
		(0.015)	(0.013)	(0.011)	(0.015)	(0.010)	(0.010)	(0.011)	(0.014)	(0.012)	(0.012)	(0.013)
	c=-5	0.080	0.057	0.046	0.051	0.037	0.036	0.038	0.051	0.043	0.058	0.081
		(0.013)	(0.011)	(0.010)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.010)	(0.011)	(0.011)
	c=-25	0.103	0.080	0.061	0.056	0.053	0.052	0.052	0.055	0.061	0.081	0.100
		(0.013)	(0.012)	(0.011)	(0.011)	(0.009)	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)	(0.013)
T=700	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)

Table 2: Size performances of t^w and IVX-QR for $\alpha = 0$ with the nominal size 5%.

						t^w						
τ		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
T=150	c=-0.05	0.107	0.076	0.058	0.051	0.049	0.048	0.050	0.051	0.059	0.075	0.105
		(0.014)	(0.012)	(0.012)	(0.010)	(0.010)	(0.009)	(0.011)	(0.011)	(0.011)	(0.012)	(0.014)
	c=-0.1	0.108	0.075	0.058	0.052	0.050	0.049	0.050	0.053	0.058	0.077	0.110
		(0.014)	(0.012)	(0.010)	(0.010)	(0.011)	(0.009)	(0.009)	(0.009)	(0.010)	(0.012)	(0.014)
	c=-0.15	0.109	0.080	0.058	0.054	0.050	0.050	0.050	0.053	0.061	0.077	0.112
		(0.013)	(0.013)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)	(0.011)	(0.010)	(0.012)	(0.014)
	c=-0.2	0.110	0.077	0.060	0.054	0.053	0.050	0.050	0.052	0.059	0.079	0.109
		(0.015)	(0.012)	(0.012)	(0.010)	(0.010)	(0.008)	(0.009)	(0.010)	(0.010)	(0.012)	(0.014)
T=300	c=-0.05	0.089	0.069	0.055	0.050	0.049	0.049	0.050	0.050	0.055	0.067	0.085
		(0.013)	(0.010)	(0.011)	(0.011)	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.011)	(0.012)
	c=-0.1	0.089	0.067	0.057	0.054	0.050	0.051	0.051	0.054	0.058	0.068	0.088
		(0.012)	(0.012)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.012)	(0.011)	(0.011)
	c=-0.15	0.090	0.069	0.058	0.054	0.052	0.051	0.052	0.055	0.056	0.070	0.090
		(0.011)	(0.012)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.009)	(0.010)	(0.010)	(0.012)
	c=-0.2	0.086	0.071	0.056	0.053	0.051	0.050	0.050	0.052	0.058	0.070	0.088
		(0.011)	(0.011)	(0.011)	(0.009)	(0.010)	(0.008)	(0.009)	(0.011)	(0.010)	(0.012)	(0.014)
T=700	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						
τ		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
T=150	c=-0.05	0.094	0.061	0.042	0.044	0.036	0.033	0.036	0.043	0.042	0.059	0.092
		(0.014)	(0.010)	(0.009)	(0.009)	(0.008)	(0.009)	(0.007)	(0.009)	(0.010)	(0.009)	(0.013)
	c=-0.1	0.091	0.060	0.040	0.041	0.034	0.032	0.034	0.041	0.040	0.059	0.091
		(0.014)	(0.012)	(0.009)	(0.011)	(0.009)	(0.009)	(0.007)	(0.009)	(0.008)	(0.011)	(0.011)
	c=-0.15	0.092	0.060	0.041	0.040	0.034	0.032	0.033	0.040	0.040	0.058	0.091
		(0.014)	(0.011)	(0.008)	(0.010)	(0.009)	(0.008)	(0.007)	(0.008)	(0.009)	(0.012)	(0.014)
	c=-0.2	0.092	0.060	0.041	0.040	0.034	0.033	0.034	0.040	0.040	0.060	0.091
		(0.015)	(0.011)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.011)	(0.013)
T=300	c=-0.05	0.075	0.055	0.042	0.043	0.036	0.036	0.037	0.043	0.041	0.055	0.079
		(0.011)	(0.012)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.010)	(0.011)
	c=-0.1	0.077	0.058	0.043	0.042	0.037	0.036	0.037	0.043	0.043	0.056	0.080
		(0.012)	(0.010)	(0.009)	(0.010)	(0.008)	(0.007)	(0.008)	(0.009)	(0.009)	(0.010)	(0.012)
	c=-0.15	0.078	0.058	0.043	0.043	0.037	0.036	0.038	0.043	0.042	0.056	0.080
		(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.007)	(0.009)	(0.009)	(0.009)	(0.011)	(0.013)
	c=-0.2	0.078	0.058	0.042	0.043	0.038	0.036	0.039	0.043	0.043	0.059	0.081
		(0.011)	(0.011)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.011)	(0.013)
T=700	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)

Next, a comparison of the power of the proposed method with that for the IVX-QR method is made. To this end, at the nominal size 5%, Figures 1 and 2 display the results for $\alpha = 1$ and $\alpha = 0$ with different c, given

$\tau = 0.5$ and the sample size $T = 300$, while Figures 3 and 4 display the results for the lower quantile $\tau = 0.05$, and Figures 5 and 6 for the upper quantile $\tau = 0.95$. 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 Theorem 3.3 which states that the convergence rate of the newly proposed method is faster than that for the IVX-QR method. We also replicate the simulations with sample size $T = 700$, and obtain similar conclusions.

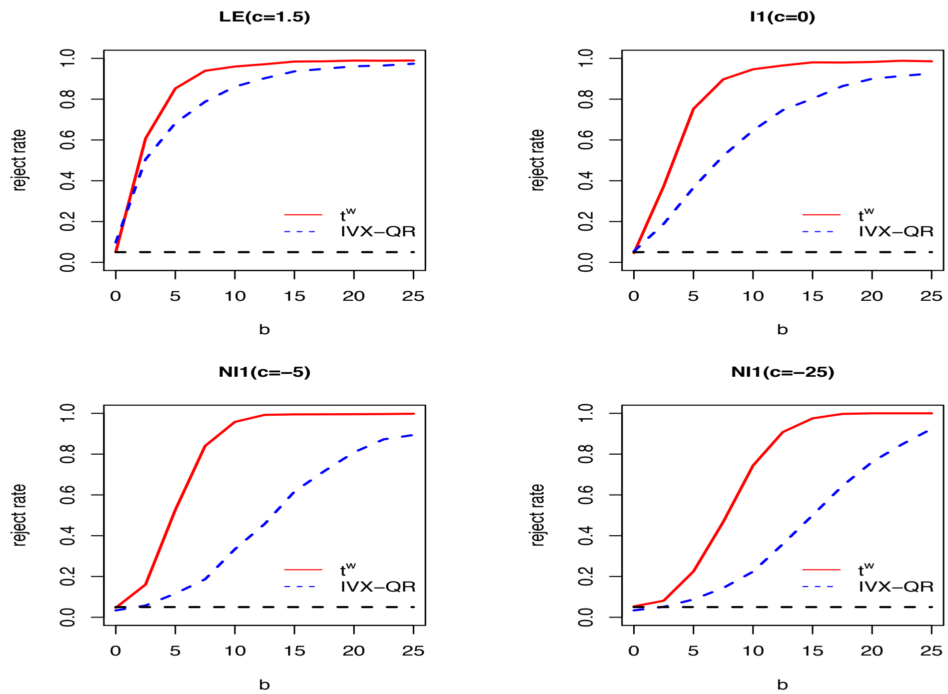

Figure 1: Local power performances of t^w and IVX-QR for $\alpha = 1$, $\beta_\tau = b_\tau/T = 3b/T$, $\tau = 0.5$ and $T = 300$.

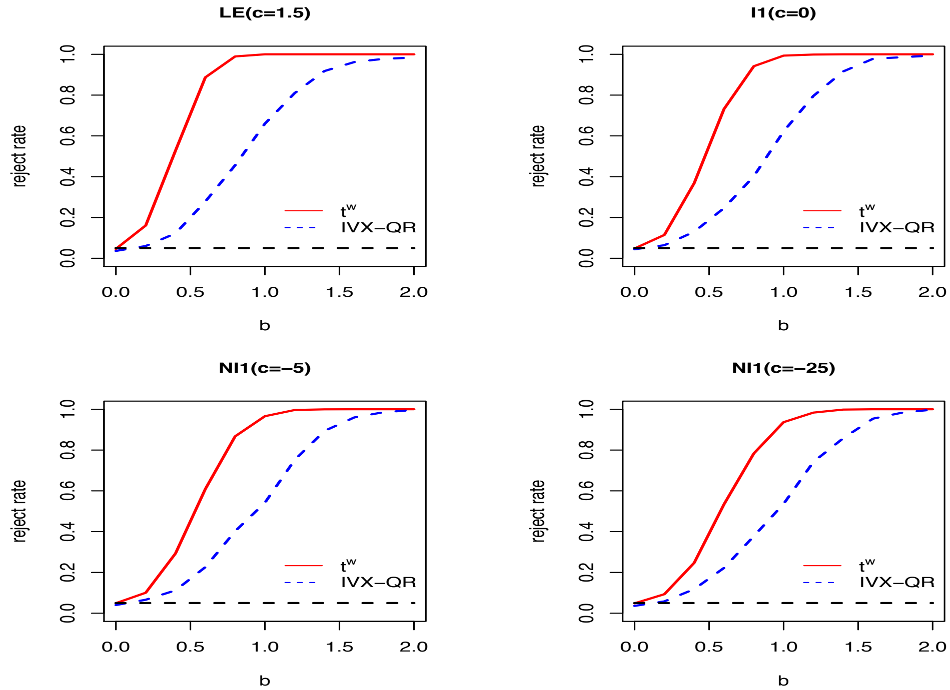

Figure 2: Local power performances of t^w and IVX-QR for $\alpha = 0$, $\beta_\tau = b_\tau/\sqrt{T} = 3b/\sqrt{T}$, $\tau = 0.5$ and $T = 300$.

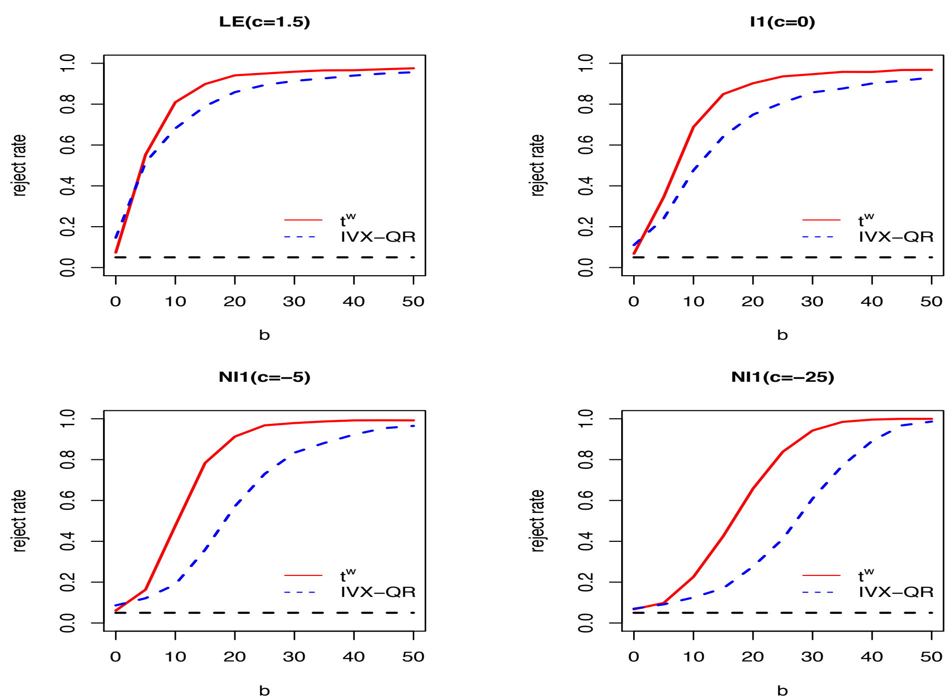

Figure 3: Local power performances of t^w and IVX-QR for $\alpha = 1$, $\beta_\tau = b_\tau/T = 1.355b/T$, $\tau = 0.05$ and $T = 300$.

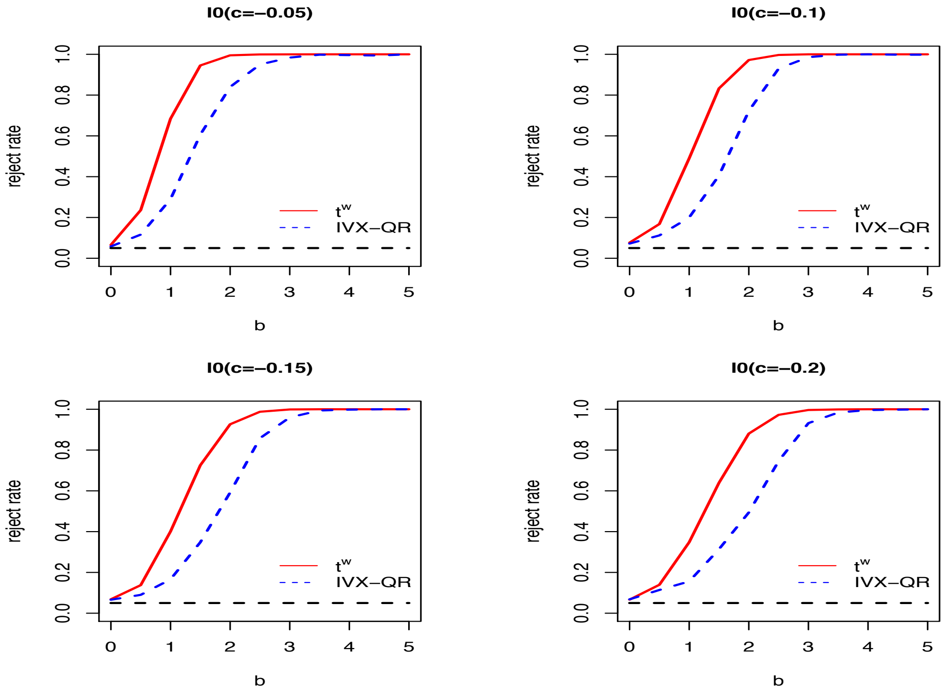

Figure 4: Local power performances of t^w and IVX-QR for $\alpha = 0$, $\beta_\tau = b_\tau/\sqrt{T} = 1.355b/\sqrt{T}$, $\tau = 0.05$ and $T = 300$.

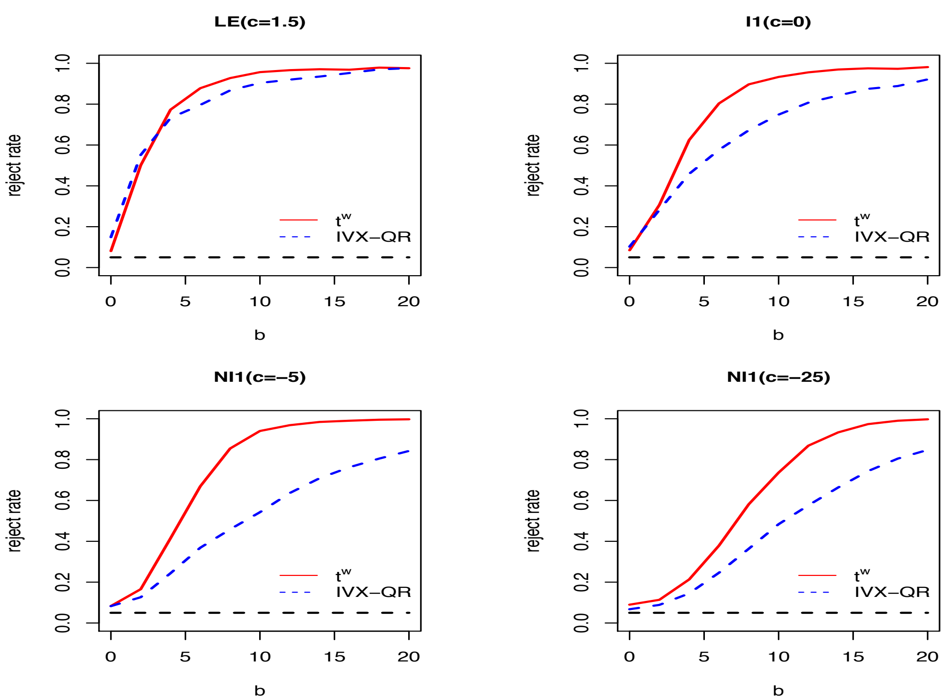

Figure 5: Local power performances of t^w and IVX-QR for $\alpha = 1$, $\beta_\tau = b_\tau/T = 4.644b/T$, $\tau = 0.95$ and $T = 300$.

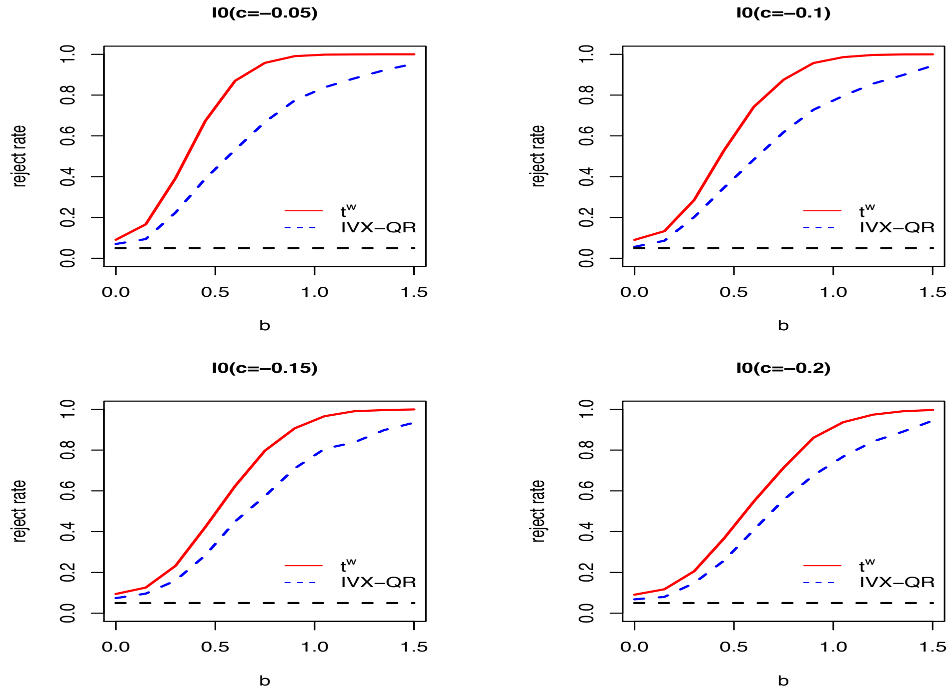

Figure 6: Local power performances of t^w and IVX-QR for $\alpha = 0$, $\beta_\tau = b_\tau/\sqrt{T} = 4.644b/\sqrt{T}$, $\tau = 0.95$ and $T = 300$.

Example 2. In this example, the model includes two predictors with different persistence types (one is NI1 and the other one is I0). The DGP is set up as follows:

$$y_t = (\mu + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1})(u_t + 3), \tag{5.1}$$

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

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

$\mu = 10$, $c_1 = -1$ and $c_2 = -0.2$. Therefore, $x_{1,t}$ is NI1 and $x_{2,t}$ is I0. The true conditional quantile of y_t given $x_{1,t-1}$ and $x_{2,t-1}$ is

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

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 τ-th quantile of u_t.

We set the sample size T as 150, 300, or 700, the nominal size as 5% and the quantile level τ as 0.05, 0.5, or 0.95. Similar to Example 1, simulation is repeated 100 times for each setting and the reject rate is computed based on 500 simulations. Table 3 shows the sizes and power performances of testing $H_0 : \beta_{1\tau} = 0$, while keeping $\beta_{2\tau} = 0$, Table 4 depicts the sizes and power performances of testing $H_0 : \beta_{2\tau} = 0$ while keeping $\beta_{1\tau} = 0$, and Table 5 displays the joint testing results for the null hypothesis $H_0 : \beta_{1\tau} = \beta_{2\tau} = 0$. The first column in Tables 3-5 (Table 3 for $\beta_1 = 0$, Table 4 for $\beta_2 = 0$, and Table 5 for $\beta_1 = 0$ and $\beta_2 = 0$) reports the empirical sizes of the proposed test, and the rest columns are for the empirical powers of the proposed test. Evidently, these tables show the proposed method works well in bivariate regression model in (5.1) which contains both stationary and nonstationary predictors. The proposed method is free of size distortion in those tests when the quantile level τ is 0.5 and also performs well in terms of power too. When our method is applied to the bivariate predictive model, it suffers from a little size distortion when the quantile level $\tau = 0.05$ or $\tau = 0.95$; however, its degree of size distortion decreases as the sample size grows. In summary, the proposed method works reasonably well in both univariate and bivariate predictive quantile models. Therefore, when compared to existing methods in the literature, our method works reasonably well and is quite competitive.

Table 3: Test results for $\beta_{1\tau}$ for the nonstationary predictor $x_{1,t}$ with a nominal size of 5%.

β_1		0	1.5	3	4.5	6	7.5	9	10.5	12	13.5	15
T=150	$\tau = 0.05$	0.121	0.134	0.161	0.212	0.277	0.351	0.424	0.497	0.560	0.613	0.653
		(0.015)	(0.016)	(0.018)	(0.018)	(0.02)	(0.021)	(0.021)	(0.021)	(0.023)	(0.021)	(0.021)
	$\tau = 0.5$	0.048	0.110	0.292	0.478	0.642	0.748	0.827	0.875	0.907	0.930	0.945
		(0.009)	(0.013)	(0.017)	(0.023)	(0.021)	(0.017)	(0.018)	(0.015)	(0.013)	(0.012)	(0.011)
	$\tau = 0.95$	0.126	0.201	0.388	0.569	0.710	0.805	0.865	0.904	0.930	0.945	0.956
		(0.015)	(0.017)	(0.022)	(0.023)	(0.019)	(0.017)	(0.015)	(0.013)	(0.010)	(0.008)	(0.008)
T=300	$\tau = 0.05$	0.095	0.102	0.125	0.159	0.213	0.277	0.349	0.417	0.483	0.546	0.597
		(0.013)	(0.014)	(0.014)	(0.015)	(0.019)	(0.016)	(0.021)	(0.017)	(0.025)	(0.021)	(0.024)
	$\tau = 0.5$	0.048	0.112	0.280	0.479	0.638	0.754	0.833	0.882	0.916	0.936	0.953
		(0.009)	(0.013)	(0.021)	(0.023)	(0.022)	(0.018)	(0.015)	(0.015)	(0.011)	(0.011)	(0.009)
	$\tau = 0.95$	0.094	0.167	0.348	0.530	0.682	0.785	0.854	0.893	0.927	0.945	0.958
		(0.015)	(0.017)	(0.022)	(0.023)	(0.02)	(0.017)	(0.015)	(0.015)	(0.010)	(0.009)	(0.008)
T=700	$\tau = 0.05$	0.074	0.079	0.097	0.128	0.170	0.221	0.277	0.337	0.397	0.451	0.509
		(0.011)	(0.013)	(0.013)	(0.014)	(0.016)	(0.019)	(0.02)	(0.019)	(0.021)	(0.023)	(0.022)
	$\tau = 0.5$	0.047	0.109	0.272	0.455	0.613	0.731	0.810	0.868	0.906	0.931	0.948
		(0.010)	(0.013)	(0.021)	(0.023)	(0.021)	(0.02)	(0.018)	(0.015)	(0.016)	(0.011)	(0.009)
	$\tau = 0.95$	0.076	0.140	0.304	0.484	0.632	0.741	0.820	0.875	0.909	0.932	0.949
		(0.010)	(0.015)	(0.02)	(0.022)	(0.024)	(0.022)	(0.018)	(0.016)	(0.014)	(0.011)	(0.010)

Table 4: Test results for $\beta_{2\tau}$ for the stationary predictor $x_{2,t}$ with a nominal size of 5%.

β_2		0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2
T=150	τ = 0.05	0.129	0.126	0.140	0.163	0.197	0.237	0.290	0.349	0.413	0.482	0.548
		(0.015)	(0.014)	(0.016)	(0.017)	(0.016)	(0.019)	(0.021)	(0.022)	(0.023)	(0.019)	(0.021)
	τ = 0.5	0.050	0.075	0.181	0.349	0.549	0.718	0.835	0.904	0.945	0.964	0.976
		(0.011)	(0.013)	(0.017)	(0.023)	(0.022)	(0.021)	(0.017)	(0.012)	(0.01)	(0.008)	(0.007)
	τ = 0.95	0.129	0.164	0.288	0.464	0.643	0.769	0.865	0.920	0.952	0.970	0.979
		(0.014)	(0.015)	(0.02)	(0.021)	(0.02)	(0.019)	(0.016)	(0.012)	(0.011)	(0.008)	(0.007)
T=300	τ = 0.05	0.099	0.101	0.116	0.137	0.169	0.214	0.266	0.326	0.391	0.459	0.530
		(0.015)	(0.015)	(0.014)	(0.014)	(0.017)	(0.021)	(0.022)	(0.023)	(0.021)	(0.024)	(0.02)
	τ = 0.5	0.048	0.084	0.211	0.427	0.645	0.813	0.910	0.958	0.980	0.989	0.992
		(0.009)	(0.013)	(0.017)	(0.022)	(0.018)	(0.016)	(0.011)	(0.009)	(0.006)	(0.005)	(0.004)
	τ = 0.95	0.097	0.143	0.285	0.488	0.689	0.829	0.917	0.959	0.980	0.988	0.993
		(0.012)	(0.016)	(0.021)	(0.021)	(0.022)	(0.017)	(0.013)	(0.009)	(0.007)	(0.005)	(0.004)
T=700	τ = 0.05	0.077	0.083	0.095	0.118	0.152	0.197	0.245	0.313	0.376	0.448	0.523
		(0.012)	(0.013)	(0.014)	(0.017)	(0.017)	(0.017)	(0.021)	(0.019)	(0.021)	(0.022)	(0.02)
	τ = 0.5	0.048	0.091	0.246	0.489	0.726	0.882	0.956	0.985	0.993	0.996	0.998
		(0.009)	(0.013)	(0.019)	(0.02)	(0.018)	(0.014)	(0.01)	(0.005)	(0.004)	(0.002)	(0.002)
	τ = 0.95	0.078	0.129	0.288	0.509	0.725	0.870	0.951	0.981	0.992	0.996	0.998
		(0.011)	(0.017)	(0.018)	(0.02)	(0.019)	(0.014)	(0.01)	(0.007)	(0.004)	(0.003)	(0.002)

Table 5: Joint test results with a nominal size of 5%.

β_1		0	1.5	3	4.5	6	7.5	9	10.5	12	13.5	15
β_2		0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2
T=150	τ = 0.05	0.164	0.170	0.201	0.250	0.321	0.398	0.489	0.563	0.629	0.688	0.734
		(0.015)	(0.019)	(0.019)	(0.02)	(0.02)	(0.021)	(0.021)	(0.022)	(0.021)	(0.021)	(0.019)
	τ = 0.5	0.050	0.109	0.297	0.516	0.695	0.811	0.884	0.926	0.952	0.964	0.975
		(0.009)	(0.014)	(0.022)	(0.023)	(0.022)	(0.017)	(0.013)	(0.013)	(0.01)	(0.008)	(0.007)
	τ = 0.95	0.162	0.240	0.436	0.632	0.777	0.865	0.917	0.948	0.965	0.975	0.982
		(0.017)	(0.018)	(0.023)	(0.021)	(0.019)	(0.016)	(0.012)	(0.01)	(0.008)	(0.006)	(0.006)
T=300	τ = 0.05	0.116	0.126	0.147	0.195	0.252	0.331	0.413	0.496	0.570	0.642	0.706
		(0.014)	(0.015)	(0.015)	(0.018)	(0.019)	(0.02)	(0.019)	(0.022)	(0.023)	(0.024)	(0.022)
	τ = 0.5	0.047	0.111	0.314	0.550	0.744	0.866	0.930	0.962	0.979	0.988	0.991
		(0.01)	(0.012)	(0.018)	(0.023)	(0.02)	(0.015)	(0.012)	(0.008)	(0.006)	(0.006)	(0.004)
	τ = 0.95	0.118	0.198	0.408	0.627	0.788	0.886	0.941	0.967	0.981	0.989	0.992
		(0.015)	(0.019)	(0.02)	(0.022)	(0.017)	(0.014)	(0.01)	(0.008)	(0.006)	(0.005)	(0.005)
T=700	τ = 0.05	0.087	0.095	0.117	0.156	0.209	0.272	0.349	0.427	0.506	0.587	0.663
		(0.012)	(0.012)	(0.014)	(0.015)	(0.016)	(0.018)	(0.02)	(0.022)	(0.021)	(0.023)	(0.022)
	τ = 0.5	0.046	0.114	0.327	0.591	0.793	0.909	0.962	0.985	0.994	0.997	0.998
		(0.009)	(0.014)	(0.019)	(0.022)	(0.021)	(0.012)	(0.008)	(0.006)	(0.004)	(0.003)	(0.002)
	τ = 0.95	0.086	0.167	0.376	0.619	0.800	0.909	0.960	0.983	0.991	0.996	0.998
		(0.013)	(0.015)	(0.022)	(0.021)	(0.014)	(0.014)	(0.008)	(0.006)	(0.004)	(0.003)	(0.002)

6 An Empirical Application

6.1 Data

This section applies the newly proposed method to revisit the question of whether or not stock market index returns are predictable by a set of macroeconomic indicators and financial ratios. For the convenience of comparison, our main results are based on the same dataset (monthly data) in Lee (2016), with a sample period from January 1927 to December 2005. An updated dataset until December 2018 is considered too to see whether there is any change after the 2008 global crisis.[7]. 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).[8] 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 persistence for all periods, but we are still unable to identify the persistence category for each variable, see Fan and Lee (2019). Given that our new method is robust to all persistence categories, it is expected to provide more reliable conclusions than traditional approaches developed under a specific type of persistence.

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	$[0.936, 0.974]$	$[0.938, 0.975]$	$[0.925, 0.964]$	$[0.921, 0.972]$	$[0.926, 0.973]$	$[0.914, 0.961]$

6.2 Main Results

We first 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 Jan.1927 to Dec. 2005. The p-values (%) shown in bold imply the rejection of the null hypothesis of no predictablity at the 5% level. The main findings can be summarized as follows. For the group of valuation ratios, we find significant lower

[7]The updated dataset is available from the website of Professor Amit Goyal at http://www.hec.unil.ch/agoyal

[8]One may refer to Goyal and Welch (2008) for detailed constructions and economic foundations of all variables.

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) finds 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 reports a significant predictability at the 80% quantile level. Meanwhile, Lee (2016) finds 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 .

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%.

Next, we conduct the quantile prediction tests for the post-1952 data until Dec. 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.

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%.

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 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 Jan. 1927 to Dec. 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 persistence 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.

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.

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.

6.3 Test Results for the Updated Dataset

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 simplicity, we only consider the multivariate quantile regression because it avoids the risk of model misidentification of univariate case. Table 11 reports the results in multivariate predictive quantile regression using the updated sample period (1927:01-2018:12), while Table

12 for the post-1952 sample period (1952:01-2018:12). The main conclusions are roughly consistent with those using the sample period until Dec. 2005.

Table 11: p-values (%) for the test using the multivariate model for the period from (1927:01-2018:12)

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	**0.7**	**0.7**	**2.8**	11.3	29.5	42.1	16.2	**1.2**	**0.3**	**0.6**	**0.5**
tbl	**0.1**	**0.1**	**2.0**	10.5	16.0	14.9	16.7	8.8	**1.8**	**0.2**	**0.4**
Joint Test	**0.0**	**0.0**	**0.8**	10.5	28.0	24.2	6.8	**0.1**	**0.0**	**0.0**	**0.0**
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	**0.6**	**1.2**	**3.8**	**4.2**	15.8	17.6	34.0	20.0	15.8	7.9	10.3
tbl	**0.3**	**0.7**	**3.1**	5.6	16.7	14.7	21.2	11.5	**4.4**	**1.2**	**2.3**
dfy	**0.0**	**0.2**	**1.3**	**2.2**	12.9	17.9	44.1	21.2	**2.9**	**0.2**	**0.1**
tms	**0.1**	**0.3**	**1.1**	**1.5**	12.9	14.8	42.8	31.4	20.9	5.8	9.3
Joint Test	**0.0**	**0.0**	**0.0**	**1.4**	13.7	17.7	**4.1**	**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**	**0.2**	**1.8**	12.4	19.4	17.1	14.2	**3.4**	**0.4**	**0.1**	**0.0**
b/m	**0.2**	**0.4**	**2.6**	12.1	15.5	18.0	19.8	9.5	**2.2**	**0.2**	**0.1**
Joint Test	**0.0**	**0.1**	**0.6**	12.0	24.3	31.8	14.0	**0.6**	**0.0**	**0.0**	**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	**4.4**	**3.7**	12.0	19.0	18.2	25.5	31.5	16.8	7.0	**2.5**	**0.5**
d/e	**0.4**	**0.1**	**1.5**	11.9	17.0	20.2	18.1	12.0	**1.6**	**0.1**	**0.0**
Joint Test	**0.1**	**0.0**	**0.7**	10.4	25.4	32.1	14.0	**0.5**	**0.0**	**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**	**0.2**	**0.3**	12.6	13.9	18.9	15.6	**2.5**	**0.4**	**0.1**	**0.0**
b/m	**0.0**	**0.1**	**0.2**	8.6	9.5	20.0	15.5	**2.2**	**0.3**	**0.0**	**0.0**
tms	**0.0**	**0.5**	**0.7**	24.8	23.5	24.0	18.8	**3.3**	**0.4**	**0.2**	**0.0**
Joint Test	**0.0**	**0.0**	**0.0**	**4.6**	10.0	36.7	13.0	**0.2**	**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	17.3	34.9	32.3	31.3	28.5	26.4	25.5	15.3	7.1	**4.3**	5.4
e/p	9.2	16.5	6.8	**2.7**	8.4	15.9	19.7	24.9	15.9	10.4	5.5
b/m	14.2	34.9	33.1	26.5	22.7	26.2	24.1	18.9	9.7	5.7	**4.8**
ntis	**1.4**	**4.8**	7.3	**4.9**	12.3	16.2	32.0	28.1	21.7	11.3	12.1
tbl	16.6	32.6	34.2	25.4	25.0	19.7	15.9	10.5	7.6	6.4	6.8
dfy	**0.0**	**0.0**	**0.1**	**1.0**	14.1	35.2	6.4	**0.2**	**0.0**	**0.0**	**0.0**
tms	**3.0**	14.3	21.8	19.9	27.9	27.3	28.0	21.3	10.1	5.7	5.2
Joint Test	**0.0**	**0.0**	**0.0**	**0.0**	**0.2**	**5.0**	**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.

Table 12: p-values (%) for the test using the multivariate model (1952:01-2018:12)

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	15.8	28.2	43.3	32.3	37.6	36.5	33.5	22.0	36.0	11.4	13.9
tbl	20.2	23.7	26.9	12.8	14.0	17.7	25.4	28.6	31.1	9.3	7.1
Joint Test	21.1	32.7	38.9	16.5	14.6	23.2	40.0	28.0	38.3	12.0	11.5
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	15.6	22.5	21.6	15.9	19.3	15.6	20.0	17.1	16.6	11.1	8.6
tbl	36.2	25.5	12.7	7.6	11.9	6.4	13.3	13.7	21.6	11.4	10.9
dfy	**4.4**	46.6	16.0	16.2	32.5	11.5	22.6	16.8	10.4	9.9	11.1
tms	16.1	22.8	17.8	18.3	20.2	16.0	20.9	19.6	20.3	8.7	7.8
Joint Test	**4.4**	39.0	30.2	17.8	22.6	15.4	28.8	20.9	8.1	**3.8**	**0.6**
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	14.6	25.1	22.3	17.2	22.6	24.2	22.3	25.5	29.8	11.0	7.9
b/m	19.5	26.4	20.8	12.8	13.3	18.1	20.4	26.7	29.2	8.1	8.3
Joint Test	18.9	36.0	27.6	18.2	10.4	17.6	33.2	31.7	32.0	14.6	9.9
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	22.1	25.8	23.3	16.4	12.2	20.3	29.6	28.0	26.9	12.3	11.3
d/e	14.9	23.7	24.2	14.1	15.4	25.0	28.0	28.5	28.4	8.1	6.7
Joint Test	19.9	28.4	31.8	20.5	17.4	25.3	43.8	39.2	36.7	12.1	7.1
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	16.7	21.2	18.7	12.7	17.3	20.5	22.2	18.9	19.7	13.3	14.3
b/m	15.3	23.8	20.0	9.6	12.6	17.9	22.5	20.7	29.5	16.3	19.1
tms	19.8	32.1	30.5	17.6	24.2	32.0	31.9	32.7	28.7	7.7	8.5
Joint Test	23.3	39.7	37.8	12.6	6.4	23.3	45.4	33.7	21.4	5.9	**2.6**
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	17.2	28.7	18.9	16.2	15.7	11.0	9.7	9.6	8.2	9.1	7.2
e/p	8.9	12.9	9.8	23.8	24.9	29.6	32.1	21.9	17.2	9.5	5.9
b/m	13.7	25.1	22.8	22.1	19.5	18.1	18.6	16.0	10.9	10.0	5.7
ntis	12.0	29.3	16.0	30.9	37.5	40.1	36.2	25.4	22.3	16.7	13.5
tbl	21.6	30.9	22.5	13.9	14.3	10.4	8.8	9.8	16.1	18.4	11.7
dfy	22.1	19.5	20.2	32.9	31.0	21.1	14.7	8.3	**4.1**	**2.1**	**1.8**
tms	18.6	16.0	13.8	28.1	28.8	33.4	28.3	23.2	16.1	12.5	10.1
Joint Test	**1.1**	**1.9**	**1.4**	**2.0**	**3.5**	**0.5**	**0.3**	**0.4**	**0.1**	**0.1**	**0.0**

Note that p-value is in bold if it is less than or equal to the significant level 0.05, and d/e is ignored due to the multiple collinearity among d/e, d/p and e/p ratios.

7 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 variable addition 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 persistence 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 indictors, such as dfy, tms and tbl show strong evidence of significant prediction power, especially at lower and upper quantile levels.

References

Amihud, Y., Hurvich, C. and Wang, Y. (2009). Multiple-predictor regressions: hypothesis testing. *Review of Financial Studies*, **22**(1), 413-434.

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies* **20**(3), 651–707.

Bauer, D. and Maynard, A. (2012). Persistence-robust surplus-lag Granger causality testing. *Journal of Econometrics* **169**(2), 293–300.

Breitung, J. and Demetrescu, M. (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics* **187**(1), 358–375.

Cai, Z. and Wang, Y. (2014). Testing predictive regression models with nonstationary regressors. *Journal of econometrics* **178**(1), 4–14.

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association* **103**(484), 1595–1608.

Campbell, J. Y. and Thompson, S.B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* **21**(4), 1509–1531.

Campbell, J. Y. and Vuolteenaho, T. (2004). Bad Beta, Good Beta. *The American Economic Review* **94**(5), 1249-1275.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics* **81**(1), 27–60.

Cenesizoglu, T. and Timmermann, A. G. (2008). Is the distribution of stock returns predictable. *Available at SSRN 1107185.*

Chen, W. W. and Deo, R. S. (2009). Bias reduction and likelihood-based almost exactly sized hypothesis testing in predictive regressions using the restricted likelihood. *Econometric Theory* **25**(5), 1143–1179.

Chen, W. W., Deo, R. S. and Yi, Y. (2013). Uniform inference in predictive regression models.*Journal of Business & Economic Statistics* **31**(4), 525–533.

Dittmar, R. F. (2002). Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *Journal of Finance* **57**(1), 369–403.

Dolado, J. J. and Lütkepohl, H. (1996). Making Wald tests work for cointegrated VAR systems. *Econometric Reviews* **15**(4), 369–386.

Elliott, G. (2011). A control function approach for testing the usefulness of trending variables in forecast models and linear regression. *Journal of Econometrics* **164**(1), 79–91.

Fan, R. and Lee, J. H. (2019). Predictive quantile regressions under persistence and conditional heteroskedasticity. *Journal of Econometrics* **213**(1), 261–280.

Ferson, W. E., and R. W. Schadt. (1996). Measuring fund strategy and performance in changing economic conditions. *Journal of Finance* **51** (2), 425–461.

Gaglianone, W. P., Lima, L. R., Linton, O. and Smith, D. R. (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics* **29**(1), 150–160.

Gonzalo, J. and Pitarakis, J.Y. (2012). Regime-specific predictability in predictive regressions. *Journal of Business & Economic Statistics* **30**(2), 229–241.

Hall, P., Heyde, C.C., 1980. *Martingale Limit Theory and its Application*. Academic Press.

Harvey, C.R. and Siddique, A. (2000). Conditional skewness in asset pricing tests. *Journal of Finance* **55**(3), 1263–1295.

Jansson, M. and 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. *Canadian Journal of Statistics* **17**(3), 261–278.

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

Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory* **12**(5), 793–813.

Kostakis, A., Magdalinos, T. and Stamatogiannis, M.P. (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies* **28**(5), 1506–1553.

Kothari, S. P. and Shanken, J. (1997). Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis. *Journal of Financial Economics* **44**(2), 169-203.

Lamont, O. A. (1998). Earnings and Expected Returns. *Journal of Finance* **53**(5), 1563-1587.

Lee, J. H. (2016). Predictive quantile regression with persistent covariates: IVX-QR approach.*Journal of Econometrics* **192**(1), 105–118.

Lin, Z. Y. and Bai, Z. D. (2010). *Probability inequalities.* Beijing: Science Press.

Liu, X., Yang, B., Cai, Z. and Peng, L. (2019). A unified test for predictability of asset returns regardless of properties of predicting variables. *Journal of Econometrics* **208**(1), 141-159.

Magdalinos, T. and Phillips, P. C. (2009). Econometric inference in the vicinity of unity. *CoFie Working Paper (7)*, Singapore Management University.

Park, J.Y. (1990). Testing for unit roots and cointegration by variable addition. In: Fomby, T.B., Rhodes, G.F.J. (Eds.), Advances in Econometrics, vol. 8. JAI press Greenwich.

Phillips, P. C. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* **74**(3), 535–547.

Phillips, P. C. (2014). Optimal estimation of cointegrated systems with irrelevant instruments. *Journal of Econometrics* **178**(2), 210–224.

Phillips, P. C. (2015). Halbert White Jr. Memorial JFEC Lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics* **13**(3), 521–555.

Phillips, P. C. and Lee, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics* **177**(2), 250–264.

Phillips, P. C. and Lee, J. H. (2016). Robust econometric inference with mixed integrated and mildly explosive regressors. *Journal of Econometrics* **192**(2), 433–450.

Phillips, P. C., Shi, S. and Yu, J. (2015). Testing for multiple bubbles: Limit theory of real time detectors. *International Economic Review* **56**(4), 1079-1133.

Rapach, D. E., Strauss, J. K. and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies* **23**(2), 821–862.

Sekkel, R. (2011). International evidence on bond risk premia. *Journal of Banking & Finance* **35**(1), 174–181.

Toda, H. Y. and Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* **66**(1), 225–250.

Torous, W., Valkanov, R. and Yan, S. (2004). On Predicting stock returns with nearly integrated explanatory variables. *Journal of Business* **77**(4), 937–966.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* **21**(4), 1455–1508.

Yang, B., Liu, X., Peng, L. and Cai, Z. (2018). Unified tests for a dynamic predictive regression. Forthcoming in *Journal of Business & Economic Statistics* with DOI: https://doi.org/10.1080/07350015.2020.1714632.

Yang, B., Long, W., Peng, L. and Cai, Z. (2019). Testing the predictability of US housing price index returns based on an IVX-AR model. Forthcoming in *Journal of the American Statistical Association* with DOI: https://doi.org/10.1080/01621459.2019.1686392.

Xiao, Z. (2009). Quantile cointegrating regression. *Journal of Econometrics* **150**(2), 248–260.

Zhu, F., Cai, Z. and Peng, L. (2014). Predictive regressions for macroeconomic data. *Annals of Applied Statistics* **8**(1), 577–594.

Appendix: Mathematical Proofs

In this appendix, due to the limitation of space, only the brief derivations of the main results with some lemmas are offered. First, we prove the following two lemmas to show that the assumption for FCLT in (2.3) is reasonable since it is coincident with the model setting in (2.1) and Assumption 2.1.

Lemma A.1. *$\psi_\tau(u_{t\tau})$ is a martingale difference sequence with*

$$E(\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1}) = 0, \quad E(\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}) = \tau(1-\tau)$$

and $E[\psi_\tau(u_{t\tau})^4] = -3\tau^4 + 6\tau^3 - 4\tau^2 + \tau$ for all t.

Proof of Lemma A.1. Since

$$E(1(u_{t\tau} < 0)|\mathcal{F}_{t-1}) = P(u_{t\tau} < 0|\mathcal{F}_{t-1}) = P(y_t < Q_{y_t}(\tau|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}) = \tau,$$

then,

$$E(\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1}) = E(\tau - 1(u_{t\tau} < 0)|\mathcal{F}_{t-1}) = \tau - E(1(u_{t\tau} < 0)|\mathcal{F}_{t-1}) = \tau - \tau = 0$$

and

$$E(\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}) = E\left[\tau^2 - 2\tau 1(u_{t\tau} < 0) + 1(u_{t\tau} < 0)|\mathcal{F}_{t-1}\right] = \tau^2 - 2\tau^2 + \tau = \tau(1-\tau).$$

Similarly, $E[\psi_\tau(u_{t\tau})^4|\mathcal{F}_{t-1}] = -3\tau^4 + 6\tau^3 - 4\tau^2 + \tau$. Then, by the iterative law of expectation,

$$E[\psi_\tau(u_{t\tau})^4] = -3\tau^4 + 6\tau^3 - 4\tau^2 + \tau.$$

The proof is complete. ■

Lemma A.2. *Under Assumption 2.1, then, $\Sigma_{\psi_\tau v} < \infty$.*

Proof of Lemma (A.2). For $h < 0$, by the iterative law of expectation and Lemma A.1,

$$E[\psi_\tau(u_{t\tau})v_{t+h}] = E\left[E(\psi_\tau(u_{t\tau})v_{t+h}|\mathcal{F}_{t-1})\right] = E\left[v_{t+h}E(\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1})\right] = E\left[v_{t+h}\cdot 0\right] = 0.$$

Thus,

$$\Sigma_{\psi_\tau v} = \sum_{h=-\infty}^{\infty} E[\psi_\tau(u_{t\tau})v_{t+h}] = \sum_{h=0}^{\infty} E[\psi_\tau(u_{t\tau})v_{t+h}].$$

For $h \geq 0$, by the iterative law of expectation and Assumption 2.1,

$$E[\psi_\tau(u_{t\tau})v_{t+h}] = E[\psi_\tau(u_{t\tau})\sum_{j=0}^{\infty} F_{xj}\varepsilon_{t+h-j}] = \sum_{j=0}^{\infty} F_{xj}E[\psi_\tau(u_{t\tau})\varepsilon_{t+h-j}] = F_{xh}E[\psi_\tau(u_{t\tau})\varepsilon_t].$$

The last step holds since $E[\psi_\tau(u_{t\tau})\varepsilon_{t+h-j}] = E\left[\varepsilon_{t+h-j}E(\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1})\right] = E\left[\varepsilon_{t+h-j}0|\mathcal{F}_{t-1})\right] = 0$ for $0 \le h < j$ and $E[\psi_\tau(u_{t\tau})\varepsilon_{t+h-j}] = E\left[\psi_\tau(u_{t\tau})E(\varepsilon_{t+h-j}|\mathcal{F}_{t-1})\right] = E\left[\psi_\tau(u_{t\tau})0\right] = 0$ for $h > j$. Then

$$\Sigma_{\psi_\tau v} = \sum_{h=0}^{\infty} E[\psi_\tau(u_{t\tau})v_{t+h}] = \sum_{h=0}^{\infty} F_{xh}E[\psi_\tau(u_{t\tau})\varepsilon_t] = F_x(1)E[\psi_\tau(u_{t\tau})\varepsilon_t].$$

Then by Hölder's inequality, $|E[\psi_\tau(u_{t\tau})\varepsilon_t]| \le \left[E[\psi_\tau(u_{t\tau})^2]\right]^{1/2}\left[E[\varepsilon_t^2]\right]^{1/2} = \tau(1-\tau)\Sigma_\varepsilon$, and by Assumption 2.1, $|F_x(1)| < \infty$. Then,

$$|\Sigma_{\psi_\tau v}| \le |F_x(1)|\tau(1-\tau)\Sigma_\varepsilon < \infty.$$

This completes the proof. ∎

Next, we prove Theorem 3.1, i.e., the Bahadur representation theorem for nonstationary case. For this purpose, it needs to establish the following proposition and Lemmas.

Proposition A.1. Let $V_T(v)$ be a vector function that satisfies

(i) $-v^\top V_T(\lambda v) \ge -v^\top V_T(v)$ for any $\lambda \ge 1$.

(ii) $\sup_{\|v\|<M} \|V_T(v) + Dv - A_T\| = o_p(1)$ where $\|A_T\| = O_p(1)$, $0 < M < \infty$. And D is a positive-definite random matrix. Suppose that v_T is a vector such that $\|V_T(v_T)\| = o_p(1)$, then

(1) $\|v_T\| = O_p(1)$.

(2) $v_T = D^{-1}A_T + o_p(1)$.

Proof of Proposition A.1. Proposition A.1 is similar to Lemma A.1 of Cai and Xu (2008), but here the matrix D is relaxed to allow for a positive-definite random matrix. First, it shows that $\|v_T\| = O_p(1)$. Following Koenker and Zhao (1996), for any given $\varepsilon > 0$ and $\ell > 0$, one has

$$\begin{aligned}
&P(\inf_{\|v\|=M}\left[-v^\top V_T(v)\right] < \ell M) \\
&\le P(\inf_{\|v\|=M}\left[-v^\top V_T(v)\right] < \ell M, \inf_{\|v\|=M}\left[-v^\top(-Dv + A_T)\right] \ge 2\ell M) \\
&+ P(\inf_{\|v\|=M}\left[-v^\top V_T(v)\right] < \ell M, \inf_{\|v\|=M}\left[-v^\top(-Dv + A_T)\right] \le 2\ell M) \\
&\le P(\inf_{\|v\|=M}\left[-v^\top V_T(v)\right] < \ell M, \inf_{\|v\|=M}\left[-v^\top(-Dv + A_T)\right] \ge 2\ell M) \\
&+ P(\inf_{\|v\|=M}\left[-v^\top(-Dv + A_T)\right] \le 2\ell M) \\
&\le P(\sup_{\|v\|=M}\left[v^\top V_T(v)\right] > -\ell M, \inf_{\|v\|=M}\left[v^\top Dv - v^\top A_T\right] \ge 2\ell M) \\
&+ P(\inf_{\|v\|=M}\left[-v^\top(-Dv + A_T)\right] \le 2\ell M). \qquad \text{(A.1)}
\end{aligned}$$

Since $\sup_{\|v\|=M}[v^\top V_T(v)] > -\ell M, \inf_{\|v\|=M}[v^\top Dv - v^\top A_T] \ge 2\ell M$ implies $\sup_{\|v\|=M}[v^\top V_T(v)] + \inf_{\|v\|=M}[v^\top Dv - v^\top A_T] \ge$

ℓM, then,

$$P\left(\sup_{\|v\|=M}\left[v^\top V_T(v)\right] > -\ell M, \inf_{\|v\|=M}\left[v^\top Dv - v^\top A_T\right] \geq 2\ell M\right)$$
$$\leq P\left(\sup_{\|v\|=M}\left[v^\top V_T(v)\right] + \inf_{\|v\|=M}\left[v^\top Dv - v^\top A_T\right] \geq \ell M\right). \tag{A.2}$$

Define $M_S = v^\top V_T(v)$ and $M_I = v^\top Dv - v^\top A_T$. Then,

$$\begin{aligned}\sup_{\|v\|=M}\left[v^\top V_T(v)\right] + \inf_{\|v\|=M}\left[v^\top Dv - v^\top A_T\right] &= \sup_{\|v\|=M}\left[M_S + M_I - M_I\right] + \inf_{\|v\|=M}\left[M_I\right] \\ &\leq \sup_{\|v\|=M}\left[M_S + M_I\right] + \sup_{\|v\|=M}\left[-M_I\right] + \inf_{\|v\|=M}\left[M_I\right] \\ &= \sup_{\|v\|=M}\left[M_S + M_I\right] - \inf_{\|v\|=M} M_I + \inf_{\|v\|=M}\left[M_I\right] \\ &= \sup_{\|v\|=M}\left[M_S + M_I\right] \\ &= \sup_{\|v\|=M}\left[v^\top V_T(v) + v^\top Dv - v^\top A_T\right].\end{aligned}$$

Thus, $\sup_{\|v\|=M}[v^\top V_T(v)] + \inf_{\|v\|=M}[v^\top Dv - v^\top A_T] \geq \ell M$ implies $\sup_{\|v\|=M}[v^\top V_T(v) + v^\top Dv - v^\top A_T] \geq \ell M$. Therefore,

$$P\left(\sup_{\|v\|=M}\left[v^\top V_T(v)\right] + \inf_{\|v\|=M}\left[v^\top Dv - v^\top A_T\right] \geq \ell M\right)$$
$$\leq P\left(\sup_{\|v\|=M}\left[v^\top V_T(v) + v^\top Dv - v^\top A_T\right] \geq \ell M\right). \tag{A.3}$$

Moreover, $\sup_{\|v\|=M}[v^\top V_T(v) + v^\top Dv - v^\top A_T] \geq \ell M$ implies that

$$\begin{aligned}\ell &\leq \sup_{\|v\|=M}\|v^\top V_T(v) + v^\top Dv - v^\top A_T\|/M \leq \sup_{\|v\|=M}\|v^\top\| \cdot \|V_T(v) + Dv - A_T\|/M \\ &\leq \sup_{\|v\|=M}\|v^\top\|/M \sup_{\|v\|=M}\|V_T(v) + Dv - A_T\| \leq \sup_{\|v\|=M}\|V_T(v) + Dv - A_T\|.\end{aligned}$$

Then,

$$P\left(\sup_{\|v\|=M}\left[v^\top V_T(v) + v^\top Dv - v^\top A_T\right] \geq \ell M\right) \leq P\left(\ell \leq \sup_{\|v\|=M}\|V_T(v) + Dv - A_T\|\right). \tag{A.4}$$

By (A.2), (A.3) and (A.4), it follows that

$$P\left(\sup_{\|v\|=M}\left[v^\top V_T(v)\right] > -\ell M, \inf_{\|v\|=M}\left[v^\top Dv - v^\top A_T\right] \geq 2\ell M\right) \leq P\left(\ell \leq \sup_{\|v\|=M}\|V_T(v) + Dv - A_T\|\right). \tag{A.5}$$

On the other hand, $\inf_{\|v\|=M}[-v^\top(-Dv + A_T)] \leq 2\ell M$ implies that

$$2\ell M \geq \inf_{\|v\|=M}\left(v^\top Dv\right) + \inf_{\|v\|=M}\left(-v^\top A_T\right)$$

and thus,

$$\inf_{\|v\|=M}\left(v^\top Dv\right)/M^2 \leq 2\ell/M - \inf_{\|v\|=M}\left(-v^\top A_T\right)/M^2 = 2\ell/M + \sup_{\|v\|=M}\left(v^\top A_T\right)/M^2.$$

It follows that

$$\inf_{\|v\|=M}\left(v^\top Dv\right)/M^2 \le 2\ell/M + \sup_{\|v\|=M}\|v^\top A_T\|/M^2 \le 2\ell/M + \sup_{\|v\|=M}\|v^\top\|\|A_T\|/M^2 = 2\ell/M + \|A_T\|/M.$$

Since D is the positive-definite random matrix, then, $\inf_{\|v\|=M}\left(v^\top Dv\right)/M^2 > 0$, so that

$$(2\ell + \|A_T\|)\left(\inf_{\|v\|=M}\left(v^\top Dv\right)/M^2\right)^{-1} > M.$$

To sum up, $\inf_{\|v\|=M}\left[-v^\top(-Dv + A_T)\right] \le 2\ell M$ implies that $(2\ell + \|A_T\|)\left(\inf_{\|v\|=M}\left(v^\top Dv\right)/M^2\right)^{-1} > M$. Therefore,

$$P\left(\inf_{\|v\|=M}\left[-v^\top(-Dv + A_T)\right] \le 2\ell M\right) \le P\left((2\ell + \|A_T\|)\left(\inf_{\|v\|=M}\left(v^\top Dv\right)/M^2\right)^{-1} > M\right) \tag{A.6}$$

Since $\|A_T\| = O_p(1)$ and $\inf_{\|v\|=M}\left(v^\top Dv\right)/M^2 > 0$, then

$$(2\ell + \|A_T\|)\left(\inf_{\|v\|=M}\left(v^\top Dv\right)/M^2\right)^{-1} = O_p(1). \tag{A.7}$$

Thus, it follows by (A.6) and (A.7) that for large $0 < M < \infty$,

$$P\left(\inf_{\|v\|=M}\left[-v^\top(-Dv + A_T)\right] \le 2\ell M\right) \le \varepsilon/4. \tag{A.8}$$

An application of (A.1), (A.5) and (A.8) concludes that

$$P\left(\inf_{\|v\|=M}\left[-v^\top V_T(v)\right] < \ell M\right) \le P\left(\sup_{\|v\|=M}\left[V_T(v) + Dv - A_T\right] \ge \ell\right) + \varepsilon/4. \tag{A.9}$$

Moreover, it follows from Assumption (ii) that,

$$P(\sup_{\|v\|=M}\left[V_T(v) + Dv - A_T\right] \ge \ell) < \varepsilon/4. \tag{A.10}$$

By the inequality in (A.9) and the result in (A.10), it is straightforward to see that for any given $\varepsilon > 0$ and $\ell > 0$ there exist T_0 and $0 < M < \infty$ such that

$$P(\inf_{\|v\|=M}\left[-v^\top V_T(v)\right] < \ell M) \le \varepsilon/2 \tag{A.11}$$

for $T > T_0$. Next, for any v, $\|v\| \ge M$, denote $v = \lambda\tilde{v}$, where $\lambda \ge 1$ and $\|\tilde{v}\| = M$. Assumption (i) implies

$$\|V_T(v)\| \ge [-\tilde{v}V(\lambda\tilde{v})]/M \ge [-\tilde{v}V(\tilde{v})]/M.$$

Therefore,

$$P\left(\inf_{\|v\|\ge M}\|V_T(v)\| < \ell\right) \le P\left(\inf_{\|v\|=M}[-\tilde{v}V(\tilde{v})] < \ell M\right) < \varepsilon/2.$$

The last inequality holds by (A.11). For enough large T, since $\|V_T(v_T)\| = o_p(1)$, then

$$P(\|v_T\| \ge M) \le P(\|v_T\| \ge M, \|V_T(v_T)\| < \ell) + P(\|V_T(v_T)\| \ge \ell) \le P\left(\inf_{\|v\|\ge M}\|V_T(v_T)\| < \ell\right) + \varepsilon/2 \le \varepsilon.$$

Thus, we conclude that

$$\|v_T\| = O_p(1). \tag{A.12}$$

From this result and Assumption (ii), it follows that

$$\sup_{\|v\|<M} \|V_T(v) + Dv - A_T\| = o_p(1). \tag{A.13}$$

By (A.12) and (A.13),

$$V_T(v_T) + Dv_T - A_T = o_p(1).$$

Since v_T is a vector such that $\|V_T(v_T)\| = o_p(1)$, we have

$$v_T = D^{-1}A_T + o_p(1).$$

This is the end of proof. ∎

Lemma A.3. *When x_t is NI1, I1 and LE,*

$$N_T = \boldsymbol{D}_T^{-1}\sum_{t=2}^{T}\Lambda_{t-1}\Lambda_{t-1}^{\top}\boldsymbol{D}_T^{-1} = N + o_p(1), \tag{A.14}$$

and

$$\boldsymbol{D}_T^{-1}\sum_{t=2}^{T}\Lambda_{t-1}\psi_\tau(u_{t\tau}) \xrightarrow{d} \left(\int dB_{\psi_\tau}(r), \int [J_x^c(r) - \tilde{\pi}_1 B(r)]dB_{\psi_\tau}(r), \tilde{\pi}_1\int B(r)dB_{\psi_\tau}(r)\right), \tag{A.15}$$

where

$$N_T = \begin{pmatrix} 1 & \int J_x^c(r)dr - \tilde{\pi}_1\int B(r)dr & \tilde{\pi}_1\int B(r)dr \\ \int J_x^c(r)dr - \tilde{\pi}_1\int B(r)dr & \int [J_x^c(r) - \tilde{\pi}_1 B(r)]^2 dr & \tilde{\pi}_1\int J_x^c(r)B(r)dr - \tilde{\pi}_1^2\int B(r)^2 dr \\ \tilde{\pi}_1\int B(r)dr & \tilde{\pi}_1\int J_x^c(r)B(r)dr - \tilde{\pi}_1^2\int B(r)^2 dr & \tilde{\pi}_1^2\int B(r)^2 dr. \end{pmatrix}.$$

Proof of Lemma A.3. Since $x_{\lfloor rT\rfloor}/\sqrt{T} \Rightarrow J_x^c(r)$ and $z_{\lfloor rT\rfloor}/\sqrt{T} = \hat{\pi}_1\zeta_{t-1}/\sqrt{T} \Rightarrow B(r)\tilde{\pi}_1$, then, by the continuous mapping theorem, the following convergence results hold true

$$\frac{1}{T^2}\sum_{t=2}^{T}x_{t-1}^{*}z_{t-1} = \frac{1}{T^2}\sum_{t=2}^{T}x_{t-1}z_{t-1} - \frac{1}{T^2}\sum_{t=2}^{T}z_{t-1}^2 \xrightarrow{d} \tilde{\pi}_1\int J_x^c(r)B(r)dr - \tilde{\pi}_1^2\int B(r)^2 dr, \tag{A.16}$$

$$\frac{1}{T^{3/2}}\sum_{t=2}^{T}z_{t-1} \xrightarrow{d} \tilde{\pi}_1\int B(r)dr, \tag{A.17}$$

$$\frac{1}{T^{3/2}}\sum_{t=2}^{T}x_{t-1}^{*} = \frac{1}{T^{3/2}}\sum_{t=2}^{T}x_{t-1} - \frac{1}{T^{3/2}}\sum_{t=2}^{T}z_{t-1} \xrightarrow{d} \int J_x^c(r)dr - \tilde{\pi}_1\int B(r)dr, \tag{A.18}$$

$$\frac{1}{T^{3/2}}\sum_{t=2}^{T}(x_{t-1}^{*})^2 \xrightarrow{d} \int [J_x^c(r) - \tilde{\pi}_1 B(r)]^2 dr, \tag{A.19}$$

$$\frac{1}{T^2}\sum_{t=2}^{T}z_{t-1}^2 \xrightarrow{d} \tilde{\pi}_1^2\int B(r)^2 dr. \tag{A.20}$$

Thus, (A.14) holds. For (A.15), it is similar to show the followings

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\psi_\tau(u_{t\tau}) \xrightarrow{d} \mathrm{N}(0, \tau(1-\tau)),$$

and

$$\frac{1}{T}\sum_{t=1}^{T} z_{t-1}\psi_\tau(u_{t\tau}) \xrightarrow{d} \tilde{\pi}_1 \int B(r)_\perp dB_{\psi_\tau}(r) = \mathrm{MN}(0, \tau(1-\tau)\tilde{\pi}_1^2 \int B(r)^2 dr).$$

Here, the above step is guaranteed by independence between ζ_{t-1} and $u_{t\tau}$. Therefore, (A.15) holds. The ends the proof. ∎

Lemma A.4. *For any vector $v = O_p(1)$ with dimension $2K+1$, $\sum_{t=2}^{T} \sqrt{T}\|\boldsymbol{D}_T^{-1}\Lambda_{t-1}\| v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1} v = O_p(1)$ always holds.*

Proof of Lemma A.4. If x_t is NI1, I1 and LE, by (A.16) – (A.20), the following statements holds ture

$$v^\top \sqrt{T}\boldsymbol{D}_T^{-1}\Lambda_{\lfloor rT\rfloor}\Lambda_{\lfloor rT\rfloor}^\top \boldsymbol{D}_T^{-1}\sqrt{T} v \;\Rightarrow\; v^\top N_v v, \tag{A.21}$$

where

$$N_v = \begin{pmatrix} 1 & J_x^c(r)dr - \tilde{\pi}_1 B(r) & \tilde{\pi}_1 B(r) \\ J_x^c(r)\tilde{\pi}_1 B(r) & [J_x^c(r) - \tilde{\pi}_1 B(r)]^2 & \tilde{\pi}_1 J_x^c(r) B(r) - \tilde{\pi}_1^2 B(r)^2 \\ \tilde{\pi}_1 B(r) & \tilde{\pi}_1 J_x^c(r) B(r) - \tilde{\pi}_1^2 B(r)^2 & \tilde{\pi}_1^2 B(r)^2, \end{pmatrix},$$

and

$$\sqrt{T}\|\boldsymbol{D}_T^{-1}\Lambda_{\lfloor rT\rfloor}\| \Rightarrow \sqrt{1 + (J_x(r) - B(r))^2 + B(r)^2}.$$

Therefore,

$$\begin{aligned}
&\sum_{t=2}^{T}\sqrt{T}\|\boldsymbol{D}_T^{-1}\Lambda_{t-1}\| v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1} v = \sum_{t=2}^{T}\sqrt{T}\|\boldsymbol{D}_T^{-1}\Lambda_{t-1}\| v^\top \sqrt{T}\boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \sqrt{T}\boldsymbol{D}_T^{-1} v \frac{1}{T} \\
&\xrightarrow{d} \int \sqrt{1 + (J_x(r) - B(r))^2 + B(r)^2} v^\top N_v v dr.
\end{aligned}$$

Thus, $\sum_{t=2}^{T}\sqrt{T}\boldsymbol{D}_T^{-1}\Lambda_{t-1}| v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1} v = O_p(1)$ with NI1, I1 and LE cases. When x_t is I0, the proof is almost the same and omitted. Therefore, this finishes the proof. ∎

Now, it is ready for proving Theorem 3.1.

Proof of Theorem 3.1. To prove Theorem 3.1, it only needs to verify that the conditions listed in Proposition A.1 hold true. The proof for I0 case is standard, so omitted. Then, it suffice to prove Theorem 3.1 for NI1, I1 and LE cases. To this end, define the convex object function as follows.

$$Z_T(v) = \sum_{t=2}^{T}\left\{\rho_\tau\left[u_{t\tau} - v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}\right] - \rho_\tau(u_{t\tau})\right\}.$$

Using the Knight identity in Knight (1989),

$$\rho_\tau(u - v) - \rho_\tau(v) = -v\psi_\tau(u) + \int_0^v \left[1(u \le l) - 1(u \le 0)\right] dl.$$

Then,

$$Z_T(v) = -\textstyle\sum_{t=2}^{T} v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}\psi_\tau(u_{t\tau}) + \sum_{t=2}^{T}\int_0^{v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}}\left[1(u_{t\tau} \le l) - 1(u_{t\tau} \le 0)\right] dl. \tag{A.22}$$

By (A.22), $Z_T(v)$ is derivable. Then, define a new object function as $V_T(v) = -\frac{\partial Z_T(v)}{\partial v}$. It is easy to prove that

$$V_T(v) = \sum_{t=2}^{T} \boldsymbol{D}_T^{-1}\Lambda_{t-1}\psi_\tau(u_{t\tau}) - \sum_{t=2}^{T} \boldsymbol{D}_T^{-1}\Lambda_{t-1}\left[1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right]. \tag{A.23}$$

The next step is to prove that $V_T(v)$ satisfies Condition (i) of Proposition A.1. Since $1(u \le x) - 1(u < 0)$ is the non-decreasing function of x, then, $1(u_{t\tau} \le \lambda v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)$ is non-decreasing function of λ if $v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1} > 0$. Therefore,

$$v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}\left[1(u_{t\tau} \le \lambda v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right]$$

is always non-decreasing function of λ. Similarly, $1(u_{t\tau} \le \lambda v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)$ is decreasing function of λ if $v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1} < 0$. Hence,

$$v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}\left[1(u_{t\tau} \le \lambda v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right]$$

is always non-decreasing function of λ in this case. Thus,

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

is non-decreasing function of λ. As a result, for $\lambda \ge 1$, one has

$$-vV_T(\lambda v) \ge -vV_T(v).$$

Thus, Condition (1) in Proposition A.1 is verified.

It still needs to prove that $V_T(v)$ satisfies Condition (ii) in Proposition A.1. From (A.23),

$$\begin{aligned}
V_T(v) &= \sum_{t=2}^{T} \boldsymbol{D}_T^{-1}\Lambda_{t-1}\psi_\tau(u_{t\tau}) - \sum_{t=2}^{T} \boldsymbol{D}_T^{-1}\Lambda_{t-1}\left[1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right] \\
&= \sum_{t=2}^{T} \boldsymbol{D}_T^{-1}\Lambda_{t-1}\psi_\tau(u_{t\tau}) - \sum_{t=2}^{T} E_{t-1}(\eta_t) - \sum_{t=2}^{T}\left[\eta_t - E_{t-1}(\eta_t)\right] \\
&= A_T - \sum_{t=2}^{T} E_{t-1}(\eta_t) - \sum_{t=2}^{T}\left[\eta_t - E_{t-1}(\eta_t)\right],
\end{aligned} \tag{A.24}$$

where $A_T = \sum_{t=2}^{T} \boldsymbol{D}_T^{-1}\Lambda_{t-1}\psi_\tau(u_{t\tau})$ and $\eta_t = \boldsymbol{D}_T^{-1}\Lambda_{t-1}\left[1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right]$. Therefore, to verify Condition (ii) of Proposition A.1 , i.e. $\sup_{\|v\|<M}\|V_T(v) + f_{u_\tau}(0)Nv - A_T\| = o_p(1)$ for $0 < M < \infty$, it suffices to show $\sum_{t=2}^{T} E_{t-1}(\eta_t) = N_T v + o_p(1)$ and $\sum_{t=2}^{T}\left[\eta_t - E_{t-1}(\eta_t)\right] = o_p(1)$. By Taylor expansion,

$$\begin{aligned}
\sum_{t=2}^{T} E_{t-1}(\eta_t) &= \sum_{t=2}^{T} \boldsymbol{D}_T^{-1}\Lambda_{t-1}\left[F_{u_{t\tau},t-1}(v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - F_{u_{t\tau},t-1}(0)\right] \\
&= \sum_{t=2}^{T} \boldsymbol{D}_T^{-1}\Lambda_{t-1}\left[f_{u_{t\tau},t-1}(0)\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1}v + \frac{1}{2}f'_{u_{t\tau},t-1}(l^*)v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1}v\right] \\
&= \sum_{t=2}^{T} f_{u_{t\tau},t-1}(0)\boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1}v + \sum_{t=2}^{T} \boldsymbol{D}_T^{-1}\Lambda_{t-1}\frac{1}{2}f'_{u_{t\tau},t-1}(l^*)v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1}v, \\
&= B_1 + B_2,
\end{aligned} \tag{A.25}$$

where $l^* \in (0, v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1})$ if $v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1} > 0$ while $l^* \in (v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}, 0)$ if $v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1} < 0$. So, it needs to verify $B_1 = f_{u_\tau}(0)N_T v + o_p(1)$ and $B_2 = o_p(1)$. To this end, B_1 is decomposed into two parts as follows.

$$B_1 = \sum_{t=2}^{T}\left[f_{u_{t\tau},t-1}(0) - f_{u_\tau}(0)\right]\boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1} v + \sum_{t=2}^{T} f_{u_\tau}(0)\boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1} v. \tag{A.26}$$

To verify $B_1 = f_{u_\tau}(0)N_T v + o_p(1)$, it is to show that $\sum_{t=2}^{T}\left[f_{u_{t\tau},t-1}(0) - f_{u_\tau}(0)\right]\boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1} v = o_p(1)$ for any $\|v\| < M$ and $0 < M < \infty$. By Assumption 2.2 and stationarity of $f_{u_{t\tau},t-1}(0)$, it is easy to see that

$$\sup_{0\le r\le 1}\left|\frac{1}{T^{1-\delta}}\sum_{t=2}^{\lfloor rT\rfloor}\left[f_{u_{t\tau},t-1}(0) - f_{u_\tau}(0)\right]\right| = o_p(1)$$

for some $\delta > 0$ (Xiao, 2009). Moreover, from the proof of Lemma A.4, $\boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1} = O_p(1)$ by (A.21). Then, following the idea in Xiao (2009, p.258), one has

$$\begin{aligned}
&\sum_{t=2}^{T}\left[f_{u_{t\tau},t-1}(0) - f_{u_\tau}(0)\right]\boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1} \qquad (A.27)\\
&= \frac{1}{T^\delta}\sum_{t=2}^{T}\frac{1}{T^{1-\delta}}\left[f_{u_{t\tau},t-1}(0) - f_{u_\tau}(0)\right]\sqrt{T}\boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1}\sqrt{T}\\
&= \frac{1}{T^\delta}\int \sqrt{T}\boldsymbol{D}_T^{-1}\Lambda_{\lfloor rT\rfloor}\Lambda_{\lfloor rT\rfloor}^\top \boldsymbol{D}_T^{-1}\sqrt{T}d\left[\sum_{t=2}^{\lfloor rT\rfloor}\frac{1}{T^{1-\delta}}\left(f_{u_{t\tau},t-1}(0) - f_{u_\tau}(0)\right)\right]\\
&= \frac{1}{T^\delta}o_p(1) = o_p(1).
\end{aligned}$$

Then, by (A.26), (A.27) and (A.14) in Lemma A.3,

$$\begin{aligned}
B_1 &= \sum_{t=2}^{T}\left[f_{u_{t\tau},t-1}(0) - f_{u_\tau}(0)\right]\boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1} v + \sum_{t=2}^{T} f_{u_\tau}(0)\boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1} v\\
&= f_{u_\tau}(0)N_T v + o_p(1). \qquad (A.28)
\end{aligned}$$

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

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

By Part (2) of Assumption 2.2, $\sup_{x\in\mathbb{R}}|f'_{u_{t\tau},t-1}(x)| = O_p(1)$, and by Lemma A.4,

$$\sum_{t=2}^{T}\sqrt{T}\|\boldsymbol{D}_T^{-1}\Lambda_{t-1}\|v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}\Lambda_{t-1}^\top \boldsymbol{D}_T^{-1} v = O_p(1)$$

for any $\|v\| < M$, $0 < M < \infty$. Then, $\|B_2\| \le \frac{1}{\sqrt{T}}O_p(1)O_p(1) = o_p(1)$. Thus,

$$B_2 = o_p(1). \tag{A.29}$$

It yields by combining the results in (A.25), (A.28) and (A.29) that

$$\textstyle\sum_{t=2}^{T} E_{t-1}(\eta_t) = f_{u_\tau}(0)N_T v + o_p(1). \tag{A.30}$$

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

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

where

$$\begin{aligned}\eta_{1t} &= \frac{1}{\sqrt{T}}\left[1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right], \\ \eta_{2t} &= \frac{x_{t-1}^*}{T}\left[1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right],\end{aligned}$$

and

$$\eta_{3t} = \frac{z_{t-1}}{T}\left[1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right].$$

To prove $\sum_{t=2}^{T}\left[\eta_t - E_{t-1}(\eta_t)\right] = o_p(1)$, it suffices to prove that $\sum_{t=2}^{T}\left[\eta_{it} - E_{t-1}(\eta_{it})\right] = o_p(1)$, $i = 1$, 2, and 3. We take the proof for η_{1t} as an illustration, and skip the details for η_{2t} and η_{3t}. For some $2 \le t \le T$ satisfying $v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1} > 0$, $1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0) = 1(0 < u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) \in [0,1]$, one can show that

$$\begin{aligned} T \cdot E_{t-1}(\eta_{1t}^2) &= E_{t-1}\left[1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right]^2 \\ &\le E_{t-1}\left[1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - 1(u_{t\tau} \le 0)\right] \\ &= P_{t-1}(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) - P_{t-1}(u_{t\tau} \le 0) \\ &= F_{u_{t\tau}}(v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}|\mathcal{F}_{t-1}) - F_{u_{t\tau}}(0|\mathcal{F}_{t-1}) \\ &= \underset{l_t \in (0, v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1})}{f_{u_{t\tau},t-1}(l_t)} \; v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1} \\ &\le \sup_{x\in\mathbb{R}} |f_{u_{t\tau},t-1}(x)| v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1} \\ &= \sup_{x\in\mathbb{R}} |f_{u_{t\tau},t-1}(x)| \cdot |v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}|. \end{aligned} \tag{A.31}$$

The last step is holds by Taylor expansion. On the other hand, for any $2 \le t \le T$ satisfying $v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1} \le 0$, $1(u_{t\tau} \le 0) - 1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) = 1(v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1} < u_{t\tau} \le 0) \in [0,1]$, one can obtain the following:

$$\begin{aligned} T \cdot E_{t-1}(\eta_{1t}^2) &= E_{t-1}\left[-1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) + 1(u_{t\tau} \le 0)\right]^2 \\ &\le E_{t-1}\left[-1(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) + 1(u_{t\tau} \le 0)\right] \\ &= P_{t-1}(u_{t\tau} \le 0) - P_{t-1}(u_{t\tau} \le v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}) \\ &= F_{u_{t\tau}}(0|\mathcal{F}_{t-1}) - F_{u_{t\tau}}(v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}|\mathcal{F}_{t-1}) \\ &= -\underset{l_t \in (v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}, 0)}{f_{u_{t\tau},t-1}(l_t)} \; v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1} \\ &\le -\sup_{x\in\mathbb{R}} |f_{u_{t\tau},t-1}(x)| v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}| \\ &= \sup_{x\in\mathbb{R}} |f_{u_{t\tau},t-1}(x)| \cdot |v^\top \boldsymbol{D}_T^{-1}\Lambda_{t-1}|. \end{aligned} \tag{A.32}$$

Then, it follows by (A.31) and (A.32) that

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

Therefore,

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

which implies, together with Part (2) of Assumption 2.2, that $\sup_{x\in\mathbb{R}} |f'_{u_{t\tau},t-1}(x)| = O_p(1)$. Similarly,

$$\begin{aligned}
\frac{1}{\sqrt{T}} \sum_{t=2}^{T} |v^\top \boldsymbol{D}_T^{-1} \Lambda_{t-1}| &= \frac{1}{T} \sum_{t=2}^{T} \left| v_1 + v_2 \frac{x_{t-1}^*}{\sqrt{T}} + v_3 \frac{z_{t-1}}{\sqrt{T}} \right| \\
&= \int |v_1 + v_2 [J_x^c(r) - \tilde{\pi}_1 B(r)] + v_3 \tilde{\pi}_1 B(r)| \, dr + o_p(1),
\end{aligned}$$

where $v = (v_1, v_2, v_3)^\top$. Then, for any $\|v\| < M$, $0 < M < \infty$,

$$\frac{1}{\sqrt{T}} \sum_{t=2}^{T} |v^\top \boldsymbol{D}_T^{-1} \Lambda_{t-1}| = O_p(1),$$

so that

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

As a result,

$$\sum_{t=2}^{T} E_{t-1}(\eta_{1t}^2) = o_p(1). \tag{A.33}$$

By the same token, one can obtain that

$$\sum_{t=2}^{T} [E_{t-1}(\eta_{1t})]^2 = o_p(1). \tag{A.34}$$

By the fact that $[\eta_{1t} - E_{t-1}(\eta_{1t})]$ is MDS and (A.33) and (A.34), it is easy calculate that

$$\begin{aligned}
&Var\left(\sum_{t=2}^{T} [\eta_{1t} - E_{t-1}(\eta_{1t})] \right) = \sum_{t=2}^{T} Var [\eta_{1t} - E_{t-1}(\eta_{1t})] \\
&= \sum_{t=2}^{T} E [\eta_{1t} - E_{t-1}(\eta_{1t})]^2 = E \left\{ \sum_{t=2}^{T} E_{t-1} [\eta_{1t} - E_{t-1}(\eta_{1t})]^2 \right\} \\
&= E \left\{ \sum_{t=2}^{T} E_{t-1}(\eta_{1t}^2) - \sum_{t=2}^{T} [E_{t-1}(\eta_{1t})]^2 \right\} = E(o_p(1) - o_p(1)) = o_p(1).
\end{aligned}$$

Thus,

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

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,

$$\sum_{t=2}^{T} [\eta_t - E_{t-1}(\eta_t)] = o_p(1). \tag{A.35}$$

By (A.24), (A.30) and (A.35), for any $\|v\| < M$, $0 < M < \infty$,

$$V_T(v) = A_T - f_{u_\tau}(0)N_T v + o_p(1).$$

By Lemma A.3, $N_T = N + o_p(1)$. Thus,

$$V_T(v) = A_T - f_{u_\tau}(0)Nv + o_p(1).$$

Therefore, for $0 < M < \infty$,

$$\sup_{\|v\|<M} \|V_T(v) + f_{u_\tau}(0)Nv - A_T\| = o_p(1). \tag{A.36}$$

By Lemma A.3, it is straightforward to show that

$$\|A_T\| = O_p(1). \tag{A.37}$$

Since $\hat{\boldsymbol{\theta}}_\tau^a$ is the minimizer of the convex function $Z_T(v)$ by the loss function, then,

$$\|V_T(\hat{\boldsymbol{\theta}}_\tau^a)\| = \left\| -\frac{\partial Z_T(\hat{v}_T)}{\partial v} \right\| = \|0\| = 0 = o_p(1). \tag{A.38}$$

By (A.36), (A.37) and (A.38), we conclude that Condition (ii) of Proposition A.1 is verified. As so far, all conditions of Proposition A.1 are verified. Thus,

$$\hat{\boldsymbol{\theta}}_\tau^a = f_{u_\tau}(0)^{-1}N^{-1}A_T + o_p(1).$$

Now, by Lemma A.3, $N_T = N + o_p(1)$ and $A_T = O_p(1)$. Then, we have

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

Theorem 3.1 is proved. ■

To prove Theorem 3.2, we need to establish the following lemma.

Lemma A.5.

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

Proof of Lemma A.5. By the Bahadur representation in Theorem 3.1,

$$\begin{aligned}
&\begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=2}^{T} x_{t-1}^* & \frac{1}{T^{3/2}}\sum_{t=2}^{T} z_{t-1} \\ \frac{1}{T^{3/2}}\sum_{t=2}^{T} x_{t-1}^* & \frac{1}{T^2}\sum_{t=2}^{T}(x_{t-1}^*)^2 & \frac{1}{T^2}\sum_{t=2}^{T} z_{t-1}x_{t-1}^* \\ \frac{1}{T^{3/2}}\sum_{t=2}^{T} z_{t-1} & \frac{1}{T^2}\sum_{t=2}^{T} x_{t-1}^* z_{t-1} & \frac{1}{T^2}\sum_{t=2}^{T} z_{t-1}^2 \end{pmatrix}\begin{pmatrix} \sqrt{T}(\hat{\mu}_\tau - \mu_\tau) \\ T(\hat{\beta}_\tau - \beta_\tau) \\ T(\hat{\gamma}_\tau - \beta_\tau) \end{pmatrix} \\
&= f_{u_\tau}(0)^{-1}\begin{pmatrix} \sum_{t=2}^{T}\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}} \\ \sum_{t=2}^{T}\frac{x_{t-1}^*}{\sqrt{T}}\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}} \\ \sum_{t=2}^{T}\frac{z_{t-1}}{\sqrt{T}}\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}} \end{pmatrix} + o_p(1).
\end{aligned} \tag{A.39}$$

Define

$$S \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{T^{3/2}}\sum_{t=2}^T z_{t-1} & 0 & 1 \end{pmatrix}.$$

Pre-multiply S on both sides of (A.39), then,

$$S\begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=2}^T x_{t-1}^* & \frac{1}{T^{3/2}}\sum_{t=2}^T z_{t-1} \\ \frac{1}{T^{3/2}}\sum_{t=2}^T x_{t-1}^* & \frac{1}{T^2}\sum_{t=2}^T (x_{t-1}^*)^2 & \frac{1}{T^2}\sum_{t=2}^T z_{t-1}x_{t-1}^* \\ \frac{1}{T^{3/2}}\sum_{t=2}^T z_{t-1} & \frac{1}{T^2}\sum_{t=2}^T x_{t-1}^* z_{t-1} & \frac{1}{T^2}\sum_{t=2}^T z_{t-1}^2 \end{pmatrix}\begin{pmatrix} \sqrt{T}(\hat{\mu}_\tau - \mu_\tau) \\ T(\hat{\beta}_\tau - \beta_\tau) \\ T(\hat{\gamma}_\tau - \beta_\tau) \end{pmatrix}$$

$$= f_{u_\tau}(0)^{-1} S \begin{pmatrix} \sum_{t=2}^T \frac{\psi_\tau(u_{t\tau})}{\sqrt{T}} \\ \sum_{t=2}^T \frac{x_{t-1}^*}{\sqrt{T}}\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}} \\ \sum_{t=2}^T \frac{z_{t-1}}{\sqrt{T}}\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}} \end{pmatrix} + S\cdot o_p(1)$$

$$= f_{u_\tau}(0)^{-1} S \begin{pmatrix} \sum_{t=2}^T \frac{\psi_\tau(u_{t\tau})}{\sqrt{T}} \\ \sum_{t=2}^T \frac{x_{t-1}^*}{\sqrt{T}}\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}} \\ \sum_{t=2}^T \frac{z_{t-1}}{\sqrt{T}}\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}} \end{pmatrix} + o_p(1).$$

The last step holds as $S = O_p(1)$. Then, we have,

$$\begin{pmatrix} 1 & \frac{1}{T^{3/2}}\sum_{t=2}^T x_{t-1}^* & \frac{1}{T^{3/2}}\sum_{t=2}^T z_{t-1} \\ \frac{1}{T^{3/2}}\sum_{t=2}^T x_{t-1}^* & \frac{1}{T^2}\sum_{t=2}^T (x_{t-1}^*)^2 & \frac{1}{T^2}\sum_{t=2}^T z_{t-1}x_{t-1}^* \\ 0 & \frac{1}{T^2}\sum_{t=2}^T x_{t-1}^* z_{t-1} - \frac{1}{T^{3/2}}\sum_{t=2}^T z_{t-1}\frac{1}{T^{3/2}}\sum_{t=2}^T x_{t-1}^* & \frac{1}{T^2}\sum_{t=2}^T z_{t-1}^2 - \left[\frac{1}{T^{3/2}}\sum_{t=2}^T z_{t-1}\right]^2 \end{pmatrix}$$

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

The third row in the above equation is

$$\begin{aligned} &\left(\frac{1}{T^2}\sum_{t=2}^T x_{t-1}^* z_{t-1} - \frac{1}{T^{3/2}}\sum_{t=2}^T z_{t-1}\frac{1}{T^{3/2}}\sum_{t=2}^T x_{t-1}^*\right) T(\hat{\beta}_\tau - \beta_\tau) \\ &+ \left[\frac{1}{T^2}\sum_{t=2}^T z_{t-1}^2 - \left(\frac{1}{T^{3/2}}\sum_{t=2}^T z_{t-1}\right)^2\right] T(\hat{\gamma}_\tau - \beta_\tau) \\ &= f_{u_\tau}(0)^{-1}\sum_{t=2}^T\left(\frac{z_{t-1}}{\sqrt{T}} - \frac{1}{T^{3/2}}\sum_{t=2}^T z_{t-1}\right)\frac{\psi_\tau(u_{t\tau})}{\sqrt{T}} + o_p(1). \end{aligned}$$

Thus,

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

This completes the proof of the lemma. ■

Proof of Theorem 3.2. For simplicity, we only offer the proof for the NI1, I1 and LE cases, because the proof for the case I0 case is standard. 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 $x_{t-1}/\sqrt{1 + x_{t-1}^2} = O_p(1)$, one has,

$$z_{\lfloor rT\rfloor}/\sqrt{T} \Rightarrow \tilde{\pi}_1 B(r). \tag{A.40}$$

Recall that $x_{\lfloor rT \rfloor}/\sqrt{T} \Rightarrow J_x^c(r)$, $\sum_{t=2}^{\lfloor rT \rfloor} \psi_\tau(u_{t\tau})/\sqrt{T} \Rightarrow B_{\psi_\tau}(r)$, and

$$W_1 + W_2 \xrightarrow{d} \tilde{\pi}_1 \int \bar{B}(r)\bar{J}_x^c(r)dr, \tag{A.41}$$

where $B(r)$ is the standard Brownian motion, $\bar{B}(r) = B(r) - \int B(r)dr$, $\bar{J}_x^c(r) = J_x^c(r) - \int J_x^c(r)dr$, and $\tilde{\pi}_1 = \int \bar{B}(r)\bar{J}_x^c(r)dr / \int \bar{B}(r)^2 dr$. By the continuous mapping theorem, one obtains that

$$\begin{aligned}(W_1 + W_2)T(\hat{\beta}_\tau^w - \beta_\tau) &= f_{u_\tau}(0)^{-1} \sum_{t=2}^{T} \left(\frac{z_{t-1}}{\sqrt{T}} - \frac{1}{T}\sum_{t=2}^{T} \frac{z_{t-1}}{\sqrt{T}} \right) \frac{\psi_\tau(u_{t\tau})}{\sqrt{T}} + o_p(1) \\ &\xrightarrow{d} f_{u_\tau}(0)^{-1} \tilde{\pi}_1 \int \left[B(r) - \int B(r)dr \right]_{\perp} dB_{\psi_\tau}(r) \end{aligned} \tag{A.42}$$

Combining the results of (A.41) and (A.42), and using the independence between ζ_t and $u_{t\tau}$, one can show that

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

This ends the proof of the theorem. ■

Proof of Theorem 3.3. For simplicity, 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, $z_{\lfloor rT \rfloor}/\sqrt{T} \Rightarrow \tilde{\pi}_1 B(r)$ from (A.40). 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,

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

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

$$\begin{aligned}
&\boldsymbol{D}_T(\hat{\boldsymbol{\beta}}_\tau^w-\boldsymbol{\beta}_\tau) \\
&= f_{u_\tau}(0)^{-1}\left[(\boldsymbol{D}_T)^{-1}\sum_{t=2}^T\left(\boldsymbol{Z}_{t-1}-\frac{1}{T}\sum_{t=2}^T\boldsymbol{Z}_{t-1}\right)\boldsymbol{X}_{t-1}^\top(\boldsymbol{D}_T)^{-1}\right]^{-1}(\boldsymbol{D}_T)^{-1}\sum_{t=2}^T\left(\boldsymbol{Z}_{t-1}-\frac{1}{T}\sum_{t=2}^T\boldsymbol{Z}_{t-1}\right)\psi_\tau(u_{t\tau})+o_p(1).
\end{aligned} \tag{A.43}$$

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

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

For 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 $\boldsymbol{D}_T=\text{diag}(\sqrt{T},\sqrt{T})$. By the central limit theorem, it is easy to show that

$$(\boldsymbol{D}_T)^{-1}\sum_{t=2}^T\left(\boldsymbol{Z}_{t-1}-\frac{1}{T}\sum_{t=2}^T\boldsymbol{Z}_{t-1}\right)\psi_\tau(u_{t\tau})\xrightarrow{d}\mathrm{N}\left(0,\tau(1-\tau)\boldsymbol{V}_2\right), \tag{A.44}$$

where

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

Combining (A.43) and (A.44), together with the continuous mapping theorem, leads to

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

where

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

For Case 2, $K_1=1$, i.e., x_{1t} is nonstationary while x_{2t} is stationary, then,

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

and the weighting matrix $\boldsymbol{D}_T=\operatorname{diag}(T,\sqrt{T})$. Define

$$
\begin{aligned}
\hbar_{t-1} &= (\hbar_{1,t-1}, \hbar_{2,t-1})^{\top}\\
&= (\boldsymbol{D}_T)^{-1}\left[\frac{\hat{\pi}_{1,1}\zeta_{1,t-1}-\sqrt{T}\tilde{\pi}_{1,1}\int B_1(r)dr}{\tilde{\pi}_{1,1}\sqrt{\int \bar{B}_1(r)^2dr}}, \frac{x_{2,t-1}}{\sqrt{1+x_{2,t-1}^2}}-E\left(\frac{x_{2,t-1}}{\sqrt{1+x_{2,t-1}^2}}\right)\right]^{\top}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\left[\operatorname{diag}\left(\tilde{\pi}_{1,1}\sqrt{\int \bar{B}_1(r)^2dr}, 1\right)\right]^{-1}(\boldsymbol{D}_T)^{-1}\textstyle\sum_{t=2}^{T}\left(\boldsymbol{Z}_{t-1}-\frac{1}{T}\sum_{t=2}^{T}\boldsymbol{Z}_{t-1}\right)\psi_\tau(u_{t\tau})\\
&\qquad = \textstyle\sum_{t=2}^{T}\hbar_{t-1}\psi_\tau(u_{t\tau})+o_p(1).
\end{aligned}
\tag{A.45}
$$

Next, it is to verify that the Lindeberg condition for $\hbar_{t-1}\psi_\tau(u_{t\tau})$ holds true. That is, for any $\tilde{\varepsilon}>0$,

$$
\sum_{t=2}^{T}E\left[\,\|\hbar_{t-1}\psi_\tau(u_{t\tau})\|^2\ 1\left(\|\hbar_{t-1}\psi_\tau(u_{t\tau})\|>\tilde{\varepsilon}\right)\Big|\,\mathcal{F}_{t-1}\right]\xrightarrow{p}0.
\tag{A.46}
$$

Since $|\psi_\tau(u_{t\tau})|=|\tau-1(u_{t\tau}<0)|\le\tau+1(u_{t\tau}<0)\le 2$, then, $\|\hbar_{t-1}\|>\tilde{\varepsilon}/|\psi_\tau(u_{t\tau})|$, which implies that $\|\hbar_{t-1}\|>\tilde{\varepsilon}/2$, so that

$$
1\left(\|\hbar_{t-1}\psi_\tau(u_{t\tau})\|>\tilde{\varepsilon}\right)=1\left(\|\hbar_{t-1}\|>\tilde{\varepsilon}/|\psi_\tau(u_{t\tau})|\right)\le 1\left(\|\hbar_{t-1}\|>\tilde{\varepsilon}/2\right).
\tag{A.47}
$$

Since $|\hbar_{1,t-1}|\le 2/\sqrt{T}$, then, $\|\hbar_{t-1}\|>\tilde{\varepsilon}/2$, which implies that $\hbar_{1,t-1}^2>\tilde{\varepsilon}^2/4-4/T$. Then, by (A.47), one has

$$
1\left(\|\hbar_{t-1}\psi_\tau(u_{t\tau})\|>\tilde{\varepsilon}\right)\le 1\left(\|\hbar_{t-1}\|>\tilde{\varepsilon}/2\right)\le 1\left(\hbar_{1,t-1}^2>\tilde{\varepsilon}^2/4-4/T\right).
\tag{A.48}
$$

It follows by the facts that $|\psi_\tau(u_{t\tau})| \le 2$ and $|\hbar_{1,t-1}| \le 2/\sqrt{T}$ and (A.48) that

$$
\begin{aligned}
&\sum_{t=2}^{T} E\left[\|\hbar_{t-1}\psi_\tau(u_{t\tau})\|^2 \ 1\left(\|\hbar_{t-1}\psi_\tau(u_{t\tau})\| > \tilde{\varepsilon}\right) \Big| \mathcal{F}_{t-1}\right] \\
&\le \sum_{t=2}^{T} 4E\left[\|\hbar_{t-1}\|^2 \ 1\left(\hbar_{1,t-1}^2 > \tilde{\varepsilon}^2/4 - 4/T\right) \Big| \mathcal{F}_{t-1}\right] \\
&\le 4\sum_{t=2}^{T} E\left[\left(\hbar_{1,t-1}^2 + 4/T\right) \ 1\left(\hbar_{1,t-1}^2 > \tilde{\varepsilon}^2/4 - 4/T\right) \Big| \mathcal{F}_{t-1}\right] \\
&= 4\sum_{t=2}^{T} E\left[\hbar_{1,t-1}^2 \ 1\left(\hbar_{1,t-1}^2 > \tilde{\varepsilon}^2/4 - 4/T\right) \Big| \mathcal{F}_{t-1}\right] + \frac{16}{T}\sum_{t=2}^{T} E\left[1\left(\hbar_{1,t-1}^2 > \tilde{\varepsilon}^2/4 - 4/T\right) \Big| \mathcal{F}_{t-1}\right] \\
&= 4\sum_{t=2}^{T} E\left[\hbar_{1,t-1}^2 \ 1\left(|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right) \Big| \mathcal{F}_{t-1}\right] + \frac{16}{T}\sum_{t=2}^{T} E\left[1\left(|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right) \Big| \mathcal{F}_{t-1}\right] \\
&= 4\sum_{t=2}^{T} E\left[\hbar_{1,t-1}^2 \ 1\left(|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right)\right] + \frac{16}{T}\sum_{t=2}^{T} E\left[1\left(|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right)\right] \qquad \text{(A.49)}
\end{aligned}
$$

for $T > 4/\tilde{\varepsilon}$. Since $\zeta_{1,t-1}$ and $\mathcal{F}_{t-1}$ are independent, then, $\hbar_{1,t-1}$ and $\mathcal{F}_{t-1}$ are independent too, so that the last step in (A.49) holds. By Chebychev's Inequality,

$$
\begin{aligned}
&\sum_{t=2}^{T} E\left[1\left(|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right)\right] \\
&= \sum_{t=2}^{T} P\left(|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right) \\
&\le \sum_{t=2}^{T} E\left(\hbar_{1,t-1}^2\right) / \left(\tilde{\varepsilon}^2/4 - 4/T\right) \\
&= E\left[\frac{1}{T^2}\sum_{t=2}^{T}\left(\hat{\pi}_{1,1}\zeta_{1,t-1} - \sqrt{T}\tilde{\pi}_{1,1}\int B_1(r)dr\right)^2 / \left(\tilde{\pi}_{1,1}^2 \int \bar{B}_1(r)^2 dr\right)\right] / \left(\tilde{\varepsilon}^2/4 - 4/T\right) \\
&= E\left[\int\left(\tilde{\pi}_{1,1}B_1(r) - \tilde{\pi}_{1,1}\int B_1(r)dr\right)^2 dr / \left(\tilde{\pi}_{1,1}^2 \int \bar{B}_1(r)^2 dr\right) + o_p(1)\right] / \left(\tilde{\varepsilon}^2/4 - 4/T\right) \\
&= E\left[\int \bar{B}_1(r)^2 dr / \int \bar{B}_1(r)^2 dr + o_p(1)\right] / \left(\tilde{\varepsilon}^2/4 - 4/T\right) \\
&= 4/\tilde{\varepsilon}^2 + o(1), \qquad \text{(A.50)}
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{t=2}^{T} E\left[\hbar_{1,t-1}^2 \ 1\left(|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right)\right] \\
&= E\left[\sum_{t=2}^{T} \hbar_{1,t-1}^2 \ 1\left(|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right)\right] \\
&\le E\left[1\left(\max_{2\le t\le T}|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right)\sum_{t=2}^{T}\hbar_{1,t-1}^2\right] \qquad \text{(A.51)} \\
&= E\left[1\left(\max_{2\le t\le T}|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right)(1 + o_p(1))\right] \\
&= E\left[1\left(\max_{2\le t\le T}|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right)\right] + o(1) = P\left(\max_{2\le t\le T}|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}\right) + o(1), \qquad \text{(A.52)}
\end{aligned}
$$

where the inequality in (A.51) holds since $|\hbar_{1,t-1}| > \sqrt{\tilde{\varepsilon}^2/4 - 4/T}$, which implies that $\max_{2\le t\le T}|\hbar_{1,t-1}| >$

$\sqrt{\tilde{\varepsilon}^2/4-4/T}$. Now,

$$\begin{aligned}
&P\left(\max_{2\le t\le T}|\hbar_{1,t-1}|>\sqrt{\tilde{\varepsilon}^2/4-4/T}\right)\\
&=P\left[\max_{2\le t\le T}\left|\hat{\pi}_{1,1}\zeta_{1,t-1}-\sqrt{T}\tilde{\pi}_{1,1}\int B_1(r)dr\right|/\left(T|\tilde{\pi}_{1,1}|\sqrt{\int \bar{B}_1(r)^2dr}\right)>\sqrt{\tilde{\varepsilon}^2/4-4/T}\right]\\
&\le P\left[\left(|\hat{\pi}_{1,1}|\max_{2\le t\le T}|\zeta_{1,t-1}|+\left|\sqrt{T}|\tilde{\pi}_{1,1}|\int B_1(r)dr\right|\right)/\left(T|\tilde{\pi}_{1,1}|\sqrt{\int \bar{B}_1(r)^2dr}\right)>\sqrt{\tilde{\varepsilon}^2/4-4/T}\right]\\
&\le P\left[\left|\sqrt{T}|\tilde{\pi}_{1,1}|\int B_1(r)dr\right|/\left(T|\tilde{\pi}_{1,1}|\sqrt{\int \bar{B}_1(r)^2dr}\right)>\sqrt{\tilde{\varepsilon}^2/4-4/T}\right]\\
&+P\left[|\hat{\pi}_{1,1}|\max_{2\le t\le T}|\zeta_{1,t-1}|/\left(T|\tilde{\pi}_{1,1}|\sqrt{\int \bar{B}_1(r)^2dr}\right)>\sqrt{\tilde{\varepsilon}^2/4-4/T}\right].
\end{aligned} \tag{A.53}$$

Then, it is straightforward to show that since $|\tilde{\pi}_{1,1}\int B_1(r)dr|/\left(|\tilde{\pi}_{1,1}|\sqrt{\int \bar{B}_1(r)^2dr}\right)=|\int B_1(r)dr|/\sqrt{\int \bar{B}_1(r)^2dr}=O_p(1)$,

$$\begin{aligned}
&P\left[\left|\sqrt{T}\tilde{\pi}_{1,1}\int B_1(r)dr\right|/\left(T|\tilde{\pi}_{1,1}|\sqrt{\int \bar{B}_1(r)^2dr}\right)>\sqrt{\tilde{\varepsilon}^2/4-4/T}\right]\\
&=P\left[\left|\int B_1(r)dr\right|/\left(\sqrt{\int \bar{B}_1(r)^2dr}\right)>\sqrt{T\tilde{\varepsilon}^2/4-4}\right]\to 0,
\end{aligned} \tag{A.54}$$

and

$$\begin{aligned}
&P\left[|\hat{\pi}_{1,1}|\max_{2\le t\le T}|\zeta_{1,t-1}|/\left(T|\tilde{\pi}_{1,1}|\sqrt{\int \bar{B}_1(r)^2dr}\right)>\sqrt{\tilde{\varepsilon}^2/4-4/T}\right]\\
&\le P\left[\max_{2\le t\le T}|\zeta_{1,t-1}|>T^{0.75}\sqrt{\tilde{\varepsilon}^2/4-4/T}\right]\\
&+P\left[|\hat{\pi}_{1,1}|/\left(|\tilde{\pi}_{1,1}|\sqrt{\int \bar{B}_1(r)^2dr}\right)>T^{0.25}\sqrt{\tilde{\varepsilon}^2/4-4/T}\right].
\end{aligned} \tag{A.55}$$

Then,

$$P\left[|\hat{\pi}_{1,1}|/\left(|\tilde{\pi}_{1,1}|\sqrt{\int \bar{B}_1(r)^2dr}\right)>T^{0.25}\sqrt{\tilde{\varepsilon}^2/4-4/T}\right]\to 0 \tag{A.56}$$

since $|\hat{\pi}_{1,1}|/\left(|\tilde{\pi}_{1,1}|\sqrt{\int \bar{B}_1(r)^2dr}\right)=\left(\sqrt{\int \bar{B}_1(r)^2dr}\right)^{-1}+o_p(1)=O_p(1)$. Thus, it follows by the fact that $\zeta_{i,t-1}=\sum_{s=1}^{t-1}\varsigma_{i,s}$ and $(\varsigma_{1,s})_{t=1}^{t-1}$ is independent random variables and the generalized Kolmogorov inequality (see Section 6.4 in Lin and Bai (2010)) that

$$P\left[\max_{2\le t\le T}|\zeta_{1,t-1}|>T^{0.75}\sqrt{\tilde{\varepsilon}^2/4-4/T}\right]\le E\left(\zeta_{1,T-1}^2\right)/\left[T^{0.75}\sqrt{\tilde{\varepsilon}^2/4-4/T}\right]^2=(T-1)/\left[T^{1.5}(\tilde{\varepsilon}^2/4-4/T)\right]\to 0. \tag{A.57}$$

It is easy to show by (A.55), (A.56) and (A.57) that

$$P\left[|\hat{\pi}_{1,1}|\max_{2\le t\le T}|\zeta_{1,t-1}|/\left(T|\tilde{\pi}_{1,1}|\sqrt{\int \bar{B}_1(r)^2dr}\right)>\sqrt{\tilde{\varepsilon}^2/4-4/T}\right]\to 0. \tag{A.58}$$

Clearly, an application of (A.53), (A.54) and (A.58) implies that

$$P\left(\max_{2\le t\le T}|\hbar_{1,t-1}|>\sqrt{\tilde{\varepsilon}^2/4-4/T}\right)\to 0. \tag{A.59}$$

Therefore, a combination of (A.52) and (A.59) leads to

$$\sum_{t=2}^{T}E\left[\hbar_{1,t-1}^2\;1\left(|\hbar_{1,t-1}|>\sqrt{\tilde{\varepsilon}^2/4-4/T}\right)\right]\to 0. \tag{A.60}$$

Hence, by an application of (A.49), (A.50) and (A.60), the following statement holds true. For any $\tilde{\varepsilon}>0$,

$$\sum_{t=2}^{T}E\left[\,\|\hbar_{t-1}\psi_\tau(u_{t\tau})\|^2\;1\left(\|\hbar_{t-1}\psi_\tau(u_{t\tau})\|>\tilde{\varepsilon}\right)\Big|\mathcal{F}_{t-1}\right]\xrightarrow{p}0. \tag{A.61}$$

Thus, the Lindeberg condition for $\hbar_{t-1}\psi_\tau(u_{t\tau})$ is verified.

Next, we show the asymptotic variance of $\sum_{t=2}^{T}\hbar_{t-1}\psi_\tau(u_{t\tau})$. First, since $\hbar_{1,t-1}$ is independent of $\mathcal{F}_{t-1}$ and $\psi_\tau(u_{t\tau})$, then,

$$\begin{aligned}\sum_{t=2}^{T}E\left[\hbar_{1,t-1}^2\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}\right]&=\sum_{t=2}^{T}E\left[\hbar_{1,t-1}^2|\mathcal{F}_{t-1}\right]E\left[\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}\right]\\&=(1-\tau)\tau\sum_{t=2}^{T}E\left[\hbar_{1,t-1}^2|\mathcal{F}_{t-1}\right]\\&=(1-\tau)\tau\sum_{t=2}^{T}E\left(\hbar_{1,t-1}^2\right)\\&=(1-\tau)\tau E\left(\sum_{t=2}^{T}\hbar_{1,t-1}^2\right)=(1-\tau)\tau E\left(1+o_p(1)\right)\xrightarrow{p}(1-\tau)\tau,\end{aligned} \tag{A.62}$$

so that

$$\begin{aligned}\sum_{t=2}^{T}E\left[\hbar_{2,t-1}^2\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}\right]&=\sum_{t=2}^{T}\hbar_{2,t-1}^2E\left[\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}\right]\\&=(1-\tau)\tau\sum_{t=2}^{T}\hbar_{2,t-1}^2\xrightarrow{p}(1-\tau)\tau\,\mathrm{Var}\left(x_{2,t-1}/\sqrt{1+x_{2,t-1}}\right).\end{aligned} \tag{A.63}$$

Furthermore,

$$\begin{aligned}\sum_{t=2}^{T}E\left[\hbar_{1,t-1}\hbar_{2,t-1}\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}\right]&=\sum_{t=2}^{T}\hbar_{2,t-1}E\left[\hbar_{1,t-1}|\mathcal{F}_{t-1}\right]E\left[\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}\right]\\&=(1-\tau)\tau\sum_{t=2}^{T}\hbar_{2,t-1}E\left(\hbar_{1,t-1}\right)\\&=(1-\tau)\tau\sum_{t=2}^{T}\hbar_{2,t-1}\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]+(1-\tau)\tau\sum_{t=2}^{T}\hbar_{2,t-1}\hbar_{1,t-1}.\end{aligned} \tag{A.64}$$

By the independence between $h_{1,t-1}$ and $(h_{2,t-1}, \mathcal{F}_{t-2})^{\intercal}$, we have

$$
\begin{aligned}
E\left\{\hbar_{2,t-1}\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]|\mathcal{F}_{t-2}\right\} &= E\left(\hbar_{2,t-1}|\mathcal{F}_{t-2}\right)E\left\{\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]|\mathcal{F}_{t-2}\right\} \\
&= E\left(\hbar_{2,t-1}|\mathcal{F}_{t-2}\right)E\left\{\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]|\mathcal{F}_{t-2}\right\} \\
&= E\left(\hbar_{2,t-1}|\mathcal{F}_{t-2}\right)E\left\{\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]\right\} \\
&= E\left(\hbar_{2,t-1}|\mathcal{F}_{t-2}\right)\left[E\left(\hbar_{1,t-1}\right)-E\left(\hbar_{1,t-1}\right)\right] \\
&= E\left(\hbar_{2,t-1}|\mathcal{F}_{t-2}\right)0 = 0.
\end{aligned}
$$

That is, $\{\hbar_{2,t-1}\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]\}_{t=2}^{T}$ is martingale difference sequence. Therefore,

$$
\begin{aligned}
& Var\left\{\sum_{t=2}^{T}\hbar_{2,t-1}\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]\right\} \\
&= \sum_{t=2}^{T} Var\left\{\hbar_{2,t-1}\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]\right\} \\
&= \sum_{t=2}^{T} E\left\{\hbar_{2,t-1}\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]\right\}^2 \\
&= \sum_{t=2}^{T} E\left(\hbar_{2,t-1}^2\right)E\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]^2 \\
&= \frac{1}{T}E\left(T\hbar_{2,t-1}^2\right)\sum_{t=2}^{T} E\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]^2 \\
&= \frac{1}{T}Var\left(x_{2,t-1}/\sqrt{1+x_{2,t-1}^2}\right)\left[\int E\left[E\left(H_1(r)\right)-H_1(r)\right]^2 dr + o_p(1)\right] \\
&= o_p(1),
\end{aligned}
$$

where $H_1(r) = (B_1(r) - \int B_1(r)dr)\left(\sqrt{\int \bar{B}_1(r)^2 dr}\right)^{-1}$. Therefore

$$
\sum_{t=2}^{T}\hbar_{2,t-1}\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right] \xrightarrow{p} E\left\{\sum_{t=2}^{T}\hbar_{2,t-1}\left[E\left(\hbar_{1,t-1}\right)-\hbar_{1,t-1}\right]\right\} = 0. \tag{A.65}
$$

Moreover,

$$
\begin{aligned}
&\sum_{t=2}^{T}\hbar_{2,t-1}\hbar_{1,t-1}\\
&=\frac{1}{T}\sum_{t=2}^{T}\hbar_{2,t-1}\left(\hat{\pi}_{1,1}\zeta_{1,t-1}-\sqrt{T}\tilde{\pi}_{1,1}\int B_1(r)dr\right)\left(\tilde{\pi}_{1,1}\sqrt{\int\bar{B}_1(r)^2dr}\right)^{-1}\\
&=\frac{1}{T}\sum_{t=2}^{T}\hbar_{2,t-1}\left(\zeta_{1,t-1}-\sqrt{T}\int B_1(r)dr\right)\left(\int\bar{B}_1(r)^2dr\right)^{-1/2}+o_p(1)\\
&=\left(\int\bar{B}_1(r)^2dr\right)^{-1/2}\frac{1}{T}\sum_{t=2}^{T}\hbar_{2,t-1}\zeta_{1,t-1}+\left(\int\bar{B}_1(r)^2dr\right)^{-1/2}\int B_1(r)dr\frac{1}{T}\sum_{t=2}^{T}\sqrt{T}\hbar_{2,t-1}+o_p(1)\\
&=\left(\int\bar{B}_1(r)^2dr\right)^{-1/2}\frac{1}{T}\sum_{t=2}^{T}\hbar_{2,t-1}\zeta_{1,t-1}+\left(\int\bar{B}_1(r)^2dr\right)^{-1/2}\int B_1(r)dr\,E\left(\sqrt{T}\hbar_{2,t-1}\right)+o_p(1)\\
&=\left(\int\bar{B}_1(r)^2dr\right)^{-1/2}\frac{1}{T}\sum_{t=2}^{T}\hbar_{2,t-1}\zeta_{1,t-1}+0+o_p(1)\\
&=\left(\int\bar{B}_1(r)^2dr\right)^{-1/2}\frac{1}{T}\sum_{t=2}^{T}\hbar_{2,t-1}\zeta_{1,t-1}+o_p(1).
\end{aligned}
\tag{A.66}
$$

Due to the equation $E(\hbar_{2,t-1}\zeta_{1,t-1}|\mathcal{F}_{t-2})=E(\hbar_{2,t-1}|\mathcal{F}_{t-2})E(\zeta_{1,t-1}|\mathcal{F}_{t-2})=E(\hbar_{2,t-1}|\mathcal{F}_{t-2})E(\zeta_{1,t-1})=E(\hbar_{2,t-1}|\mathcal{F}_{t-2})0=0$, $\hbar_{2,t-1}\zeta_{1,t-1}$ is the martingale difference sequence. So

$$
\begin{aligned}
Var\left(\sum_{t=2}^{T}\hbar_{2,t-1}\zeta_{1,t-1}\right)&=\sum_{t=2}^{T}Var\left(\hbar_{2,t-1}\zeta_{1,t-1}\right)\\
&=\sum_{t=2}^{T}E\left(\hbar_{2,t-1}\zeta_{1,t-1}\right)^2\\
&=\sum_{t=2}^{T}E\left(\hbar_{2,t-1}^2\right)E\left(\zeta_{1,t-1}^2\right)\\
&=\frac{1}{T}E\left(T\hbar_{2,t-1}^2\right)\sum_{t=2}^{T}E\left(\zeta_{1,t-1}^2\right)\\
&=\frac{1}{T}Var\left(x_{2,t-1}/\sqrt{1+x_{2,t-1}^2}\right)\left[\int E\left(H_1(r)^2\right)dr+o_p(1)\right]=o_p(1).
\end{aligned}
$$

The equation holds by the independence between $\hbar_{2,t-1}$ and $\zeta_{1,t-1}$. Therefore,

$$
\sum_{t=2}^{T}\hbar_{2,t-1}\zeta_{1,t-1}\xrightarrow{p}E\left(\sum_{t=2}^{T}\hbar_{2,t-1}\zeta_{1,t-1}\right)=\sum_{t=2}^{T}E\left(\hbar_{2,t-1}\right)E\left(\zeta_{1,t-1}\right)=0.
\tag{A.67}
$$

By equation (A.64), (A.65), (A.66) and (A.67), it follows that

$$
\sum_{t=2}^{T}E\left[\hbar_{1,t-1}\hbar_{2,t-1}\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}\right]\xrightarrow{p}0.
\tag{A.68}
$$

It follows by (A.62), (A.63) and (A.68) that

$$
\begin{aligned}
\sum_{t=2}^{T}E\left[\hbar_{t-1}\hbar_{t-1}^{\intercal}\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}\right]&=\sum_{t=2}^{T}E\left[\begin{pmatrix}\hbar_{1,t-1}^2 & \hbar_{1,t-1}\hbar_{2,t-1}\\ \hbar_{1,t-1}\hbar_{2,t-1} & \hbar_{2,t-1}^2\end{pmatrix}\psi_\tau(u_{t\tau})^2|\mathcal{F}_{t-1}\right]\\
&\xrightarrow{p}\tau(1-\tau)\begin{pmatrix}1 & 0\\ 0 & \mathrm{Var}\left(x_{2,t-1}/\sqrt{1+x_{2,t-1}}\right)\end{pmatrix}
\end{aligned}
\tag{A.69}
$$

and $E\left[\hbar_{t-1}\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1}\right]$ is martingale difference sequence, since

$$
\begin{aligned}
&E\left[\hbar_{t-1}\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1}\right]\\
&=\left(E\left[\hbar_{1,t-1}\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1}\right],E\left[\hbar_{2,t-1}\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1}\right]\right)^\top\\
&=\left(E\left[\hbar_{1,t-1}|\mathcal{F}_{t-1}\right]E\left[\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1}\right],\hbar_{2,t-1}E\left[\psi_\tau(u_{t\tau})|\mathcal{F}_{t-1}\right]\right)^\top\\
&=\left(E\left[\hbar_{1,t-1}|\mathcal{F}_{t-1}\right]0,\hbar_{2,t-1}0\right)^\top=(0,0)^\top. \qquad\qquad (A.70)
\end{aligned}
$$

Therefore, it follows by (A.61), (A.69) and (A.70) and the Corollary 3.1 in Hall and Heyde (1980) that

$$
\sum_{t=2}^{T}\hbar_{t-1}\psi_\tau(u_{t\tau})\xrightarrow{d}\mathrm{N}\left[0,\tau(1-\tau)\begin{pmatrix}1 & 0\\ 0 & \mathrm{Var}\left(x_{2,t-1}/\sqrt{1+x_{2,t-1}}\right)\end{pmatrix}\right]. \qquad (A.71)
$$

It is easy to see by (A.45) and (A.71) that

$$
(\boldsymbol{D}_T)^{-1}\sum_{t=2}^{T}\left(\boldsymbol{Z}_{t-1}-\frac{1}{T}\sum_{t=2}^{T}\boldsymbol{Z}_{t-1}\right)\psi_\tau(u_{t\tau})\quad\xrightarrow{d}\mathrm{MN}\left(0,\tau(1-\tau)\boldsymbol{V}_2\right), \qquad (A.72)
$$

where

$$
\boldsymbol{V}_2=\begin{pmatrix}\tilde{\pi}_{1,1}^2\int\bar{B}_1(r)^2\,dr & 0\\ 0 & E\left(\frac{x_{2,t}^2}{1+x_{2,t}^2}\right)-E\left(\frac{x_{2,t}}{\sqrt{1+x_{2,t}^2}}\right)^2\end{pmatrix}.
$$

An application of (A.43) and (A.72) as well as the continuous mapping theorem implies that

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

where

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

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

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

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

$$
(\boldsymbol{D}_T)^{-1}\sum_{t=2}^{T}\left(\boldsymbol{Z}_{t-1}-\frac{1}{T}\sum_{t=2}^{T}\boldsymbol{Z}_{t-1}\right)\psi_\tau(u_{t\tau})\xrightarrow{d}\int\left(\tilde{\pi}_{1,1}\bar{B}_1(r),\tilde{\pi}_{1,2}\bar{B}_2(r)\right)_\perp^\top dB_{\psi_\tau}(r)=\mathrm{MN}\left(0,\tau(1-\tau)\boldsymbol{V}_2\right), \text{(A.73)}
$$

where

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

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.43) and (A.73) together with the continuous mapping theorem that

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

where

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

This concludes the proof the theorem. ∎

Proof of Theorem 4.2. By the results in Theorem 4.1, the proof of Theorem 4.2 is straightforward and the details are omitted here to save space. ∎