Predictive Quantile Regression with Persistent Covariates: IVX-QR Approach

Ji Hyung Lee*

Department of Economics
University of Washington

November, 2014

Abstract

This paper develops econometric methods for inference and prediction in quantile regression (QR) allowing for persistent predictors. Conventional QR econometric techniques lose their validity when predictors are highly persistent. I adopt and extend a methodology called IVX filtering (Magdalinos and Phillips, 2009) that is designed to handle predictor variables with various degrees of persistence. The proposed IVX-QR methods correct the distortion arising from persistent multivariate predictors while preserving discriminatory power. Simulations confirm that IVX-QR methods inherit the robust properties of QR. These methods are employed to examine the predictability of US stock returns at various quantile levels.

Keywords: IVX filtering, Local to unity, Multivariate predictors, Predictive regression, Quantile regression.

JEL classification: C22

1 Introduction

Predictive regression models are extensively used in empirical macroeconomics and finance. A leading example is stock return regression where predictability has been a long standing puzzle. A central econometric issue in these models is severe size distortion under the null arising from the presence of persistent predictors coupled with weak discriminatory power in detecting marginal levels of predictability. The predictive mean regression literature has explored and developed econometric methods for correcting this distortion and validating inference. A recent review of this research is given in Phillips and Lee (2013, Section 2).

Quantile regression (QR) has emerged as a powerful tool for estimating conditional quantiles since Koenker and Basset (1978). The method has attracted much attention in economics in view of

*Assistant Professor of Economics, University of Washington. Address: 336 Savery Hall, Box 353330, Seattle, WA98115. Phone: 206-543-4532. Email: jihyung2@uw.edu.

the importance of the entire response distribution in empirical models. Koenker (2005)'s monograph provides an excellent overview of the field. QR methods are also attractive in predictive regression because they enable practitioners to focus their attention on the quantile structure of financial asset return distribution and provide forecasts at each quantile. This focus permits significance testing of predictors of individual quantiles of asset returns. Stylized facts of financial time series data such as heavy tails and time varying volatility imply potentially greater predictability at quantiles other than the median for financial data. Standard QR econometric techniques, however, are not valid when predictors are highly persistent since predictive QR models share the same econometric issues as their mean regression counterparts.

This paper addresses these issues by developing new methods of inference for predictive QR. The limit theory of ordinary QR with persistent regressors reveals the source of the distortion to be greater under (i) stronger endogeneity, (ii) higher levels of persistence and (iii) more extreme quantiles coupled with heavy tailedness. To develop QR methods for correcting the size distortion and conducting valid inference, I adopt a recent methodology called IVX filtering developed in Magdalinos and Phillips (2009). The idea of IVX filtering is to generate an instrument of intermediate persistence by filtering a persistent and possibly endogenous regressor. The new filtered IV succeeds in correcting size distortion arising from many different forms of predictor persistence while maintaining good discriminatory power in conventional regression settings. I extend the IVX filter idea to the QR framework and propose a new approach to inference which we call IVX-QR.

The proposed IVX-QR estimator has an asymptotically mixed normal distribution in the presence of multiple persistent predictors. I develop a computationally attractive testing method for quantile predictability to simplify implementation for applied work. Employing the new methods, I examine the empirical predictability of monthly stock returns in the S&P 500 index at various quantile levels. In regressions with commonly used persistent predictors I find several quantile specific significant predictors. In particular, over the period of 1927-2005, there is significant evidence that dividend-payout ratios have predictive power for lower quantiles of stock returns, while the book-to-market value ratio is shown to predict both lower and upper quantiles of stock returns during the same period. Notably, predictability appears to be enhanced by using combinations of persistent predictors. IVX-QR corrections ensure that the quantile predictability results are not spurious even in the presence of multiple persistent predictors, suggesting the possibility of improved forecast models for stock returns. For example, the combination of the T-bill rate and the book-to-market ratio are shown to predict almost all stock return quantiles considered over the 1927-2005 period. The forecasting capability of this combination remains strong even in the post-1952 data.

Closely related to this paper are recent studies that have investigated inference in QR with financial time series. Xiao (2009) developed a limit theory of QR in the presence of unit root regressors and developed fully-modified methods based on Phillips and Hansen (1990). Cenesizoglu and Timmermann (2008) introduced the predictive QR framework and found that commonly used predictor variables affect lower, central and upper quantiles of stock returns differently. Maynard et al. (2011) examined the issue of persistent regressors in predictive QR by extending the limit theory of Xiao (2009) to a near-integrated regressor case. These last two papers can be classified as part of the predictive QR literature since they focus on the prediction of stock return quantiles from lagged financial variables. Another piece of related research is Han et al. (2014) which studies the quantile dependence between stock return and a predictor, wherein the new analysis becomes possible by extending quantilogram theory (Linton and Whang, 2007) to the cross-quantilogram. In the mean predictive regression literature, Gonzalo and Pitarakis (2012) and Kostakis et al. (2012) are close to this paper, since they have also applied the IVX methodology to the stock return regression.

The new IVX-QR methods developed in this paper contribute to the predictive QR literature in several aspects. First, the methods are uniformly valid over the extensive range of predictor persistence from stationary predictors to mildly explosive predictors. This coverage conveniently encompasses existing results for unit root (Xiao, 2009) and near unit root (Maynard et al., 2011) predictor cases. The uniform validity of the new methods allows for possible misspecification in predictor persistence. Second, the IVX-QR methods validate inference under multiple persistent predictors while most existing methods control test size with a single persistent predictor. This feature improves realism in applied work and provides potentially better forecast models since there are a variety of persistent predictors. Third, the new method corrects size distortion while preserving substantial local power in spite of the modest reduction in its convergence rate (see Section 3.1 for the detailed discussion). This advantage is critical in finding marginal levels of predictability in predictive QR with the desired size correction. IVX-QR also maintains the inherent benefits of QR such as markedly superior performance under thick-tailed errors and the capability of testing predictability at various quantile levels. All these features make the technique well suited to empirical applications in macroeconomics and finance.

The paper is organized as follows. Section 2 introduces the model and extends the limit theory of ordinary QR. Section 3 develops the new IVX-QR methods. Section 4 provides a practical rule to choose the filtering parameters. This Section also reports the simulation results based on the suggested rule. Section 5 illustrates the empirical examples and Section 6 concludes. Main proofs are given in the Appendix, while additional discussions, proofs of lemmas and more comprehensive numerical results are available from an online supplement (Lee, 2014).

2 Model Framework and Existing Problems

2.1 Model and Assumptions

I first discuss the predictive mean regression model and then explain the predictive QR model. The standard predictive mean regression model is

$$y_t = \beta_0 + \beta_1' x_{t-1} + u_{0t} \text{ with } E\left(u_{0t} | \mathcal{F}_{t-1}\right) = 0, \tag{2.1}$$

where β_1 is a $K \times 1$ vector and $\mathcal{F}_t$ is a natural filtration. A vector of predictors x_{t-1} has the following autoregressive form

$$\begin{aligned} x_t &= R_n x_{t-1} + u_{xt}, \\ R_n &= I_K + \frac{C}{n^\alpha}, \text{ for some } \alpha > 0, \end{aligned} \tag{2.2}$$

where n is the sample size and $C = diag(c_1, c_2, ..., c_K)$. The pair of (α, C) represents persistence in the multiple predictors of unknown degree. I allow for more general degrees of persistence in the predictors than in the existing literature. In particular, x_t can belong to any of the following persistence categories[1]:

(I0) stationary: $\alpha = 0$ and $|1 + c_i| < 1$, $\forall i$,

(MI) mildly integrated: $\alpha \in (0, 1)$ and $c_i \in (-\infty, 0)$, $\forall i$,

(I1) local to unity and unit root: $\alpha = 1$ and $c_i \in (-\infty, \infty)$, $\forall i$,

(ME) mildly explosive: $\alpha \in (0, 1)$ and $c_i \in (0, \infty)$, $\forall i$.

The exact degrees of persistence in economic time series are always imprecisely determined. Unit root tests do not provide a firm guidance on discrepancy between I(0), near or exact unit root processes. The extensive treatment of parameter space from (I0) to (ME)[2] in this paper helps in coping with misspecified order of integration of the multivariate predictors.

For parsimonious characterization of the parameter space the (I1) specification above includes both conventional integrated $(C = 0)$ and local to unity $(C \in (-\infty, \infty), C \neq 0)$ specifications. The innovation structure allows for linear process dependence for u_{xt} and imposes a martingale difference sequence (mds) condition for u_{0t} following convention in the predictive regression literature:

$$\begin{aligned} u_{0t} &\sim mds\,(0, \Sigma_{00}), \text{i.e., } E(u_{0t}|\mathcal{F}_{t-1}) = 0 \text{ and } E\left(u_{0t}u_{0t}'|\mathcal{F}_{t-1}\right) = \Sigma_{00}, \forall t, \\ u_{xt} &= \sum_{j=0}^{\infty} F_{xj}\varepsilon_{t-j},\ \varepsilon_t \sim mds\,(0, \Sigma),\ \Sigma > 0,\ E\,\|\varepsilon_1\|^{2+\nu} < \infty,\ \nu > 0, \\ F_{x0} &= I_K,\ \sum_{j=0}^{\infty} j\,\|F_{xj}\| < \infty,\ F_x(z) = \sum_{j=0}^{\infty} F_{xj}z^j \text{ and } F_x(1) = \sum_{j=0}^{\infty} F_{xj} > 0, \\ \Sigma_{0x} &= E\left(u_{0t}u_{xt}'\right),\ \Omega_{xx} = \sum_{h=-\infty}^{\infty} E\left(u_{xt}u_{xt-h}'\right) = F_x(1)\Sigma F_x(1)'. \end{aligned} \tag{2.3}$$

Under these conditions, the usual functional limit law holds (Phillips and Solo, 1992):

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns \rfloor} u_j := \frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor ns \rfloor} \begin{bmatrix} u_{0j} \\ u_{xj} \end{bmatrix} =: \begin{bmatrix} B_{0n}(s) \\ B_{xn}(s) \end{bmatrix} \Longrightarrow \begin{bmatrix} B_0(s) \\ B_x(s) \end{bmatrix} = BM \begin{bmatrix} \Sigma_{00} & \Sigma_{0x} \\ \Sigma_{x0} & \Omega_{xx} \end{bmatrix}, \tag{2.4}$$

[1] The different values of α's can be also allowed between (I0), (MI) and (I1) categories, see Lee (2014, Section 1.8).

[2] (MI) and (ME) spaces are introduced in Phillips and Magdalinos (2007). In Section 1.7 of Lee (2014), (MI) space is shown to conveniently encompass the stationary long memory processes.

where $B = (B_0', B_x')'$ is vector Brownian motion (BM). The local to unity limit law for case (I1) also holds (Phillips, 1987):

$$\frac{x_{\lfloor nr \rfloor}}{\sqrt{n}} \Longrightarrow J_x^c(r), \text{ where } J_x^c(r) = \int_0^r e^{(r-s)C} dB_x(s) \tag{2.5}$$

is Ornstein-Uhlenbeck (OU) process.

I now consider a linear predictive QR model. Given the natural filtration $\mathcal{F}_t = \sigma\{u_j = (u_{0j}, u_{xj}')', \, j \le t\}$, the predictive QR model is

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \beta_{0,\tau} + \beta_{1,\tau}' x_{t-1}. \tag{2.6}$$

where $Q_{y_t}(\tau|\mathcal{F}_{t-1})$ is a conditional quantile of y_t such that $\Pr(y_t \le Q_{y_t}(\tau|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}) = \tau \in (0,1)$.

The model (2.6) analyzes other quantile predictability as well as the median of y_t. This feature is well suited to the analysis of financial asset returns, whose distributional predictability will be useful for many applications. Another stylized fact of asset returns, conditional heteroskedasticity, can also be allowed in (2.6). In Section 1.3.3 of Lee (2014), (2.6) is shown to accommodate conditional heteroskedasticity by including a proxy for the conditional stock variance as one regressor.

By defining a piecewise derivative of the loss function in the QR $\psi_\tau(u) = \tau - 1(u < 0)$ (see (2.8) below), it is easy to show the QR "induced" innovation $\psi_\tau(u_{0t\tau}) \sim mds(0, \tau(1-\tau))$ where $u_{0t\tau} = u_{0t} - F_{u_0}^{-1}(\tau)$ and $F_{u_0}^{-1}(\tau)$ is the unconditional τ-quantile of u_{0t}. Then the following functional law holds

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \begin{bmatrix} \psi_\tau(u_{0t\tau}) \\ u_{xt} \end{bmatrix} \Longrightarrow \begin{bmatrix} B_{\psi_\tau}(r) \\ B_x(r) \end{bmatrix} = BM \begin{bmatrix} \tau(1-\tau) & \Sigma_{\psi_\tau x} \\ \Sigma_{x\psi_\tau} & \Omega_{xx} \end{bmatrix}. \tag{2.7}$$

This functional law drives the main asymptotics below.

Some regularity assumptions on the conditional density of $u_{0t\tau}$ are imposed.

Assumption 2.1 *(i) The sequence of stationary conditional pdf $\{f_{u_{0t\tau},t-1}(\cdot)\}$ evaluated at zero satisfies a FCLT with a non-degenerate mean $f_{u_{0\tau}}(0) = E[f_{u_{0t\tau},t-1}(0)] > 0$,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \left(f_{u_{0t\tau},t-1}(0) - f_{u_{0\tau}}(0)\right) \Longrightarrow B_{f_{u_{0\tau}}}(r).$$

(ii) For each t and $\tau \in (0,1)$, $f_{u_{0t\tau},t-1}$ is bounded above with probability one around zero, i.e., $f_{u_{0t\tau},t-1}(\epsilon) < \infty$ w.p.1 for all $|\epsilon| < \eta$ for some $\eta > 0$.

Remark 2.1 *Assumption 2.1-(i) is not restrictive considering that an mds (or iid) structure is commonly imposed on u_{0t} (hence $u_{0t\tau}$) in the predictive regression literature. Note that iid $u_{0t\tau}$ ($f_{u_{0t\tau},t-1}(0) = f_{u_{0\tau}}(0)$ for all t) is subsumed in this Assumption, where $B_{f_{u_{0\tau}}}(r)$ is identically zero in such case (degenerate Brownian motion). Time varying conditional pdf $f_{u_{0t\tau},t-1}(0)$ with weak dependence is allowed to include, for example, some empirically relevant conditionally heteroskedastic* *(but still mds) u_{0t} processes (e.g., ARCH/GARCH), see 1.3.3 of Lee (2014). Assumption 2.1-(ii) is a standard technical condition used in the QR literature, and enables the expansion of not everywhere differentiable objective functions after smoothing with the conditional pdf $f_{u_{0t\tau},t-1}$.*

2.2 Limit Theory Extension of Quantile Regression

This Section extends the existing limit theory of ordinary QR. This extension is of some independent interest and is useful in revealing the source of the problems that arise from persistent regressors in QR. The ordinary QR estimator has the form:

$$\hat{\beta}_{\tau}^{QR} = \arg\min_{\beta} \sum_{t=1}^{n} \rho_{\tau}\left(y_t - \beta' X_{t-1}\right) \tag{2.8}$$

where $\rho_{\tau}(u) = u(\tau - 1(u < 0))$, $\tau \in (0,1)$ is the asymmetric QR loss function. The notation $X_{t-1} = (1, x_{t-1}')'$ includes the intercept and the regressor x_{t-1} whose specification is given in (2.2). I employ different normalizing matrices according to the regressor persistence:

$$D_n := \begin{cases} \sqrt{n} I_{K+1} & \text{for (I0)}, \\ diag(\sqrt{n}, n^{\frac{1+\alpha}{2}} I_K) & \text{for (MI)}, \\ diag(\sqrt{n}, n I_K) & \text{for (I1)}, \\ diag(\sqrt{n}, n^{\alpha} R_n^n) & \text{for (ME)}. \end{cases} \tag{2.9}$$

Using the Convexity Lemma (Pollard, 1990), as in Xiao (2009), I prove the next theorem that encompasses the limit theory for the unit root case (Theorem 1 in Xiao; 2009), stationary local to unity case (Proposition 2 in Maynard et al; 2011) and stationary case (Koenker, 2005). This paper adds to the QR literature by extending that limit theory to the (MI) and (ME) cases.

Theorem 2.1

$$D_n\left(\hat{\beta}_{\tau}^{QR} - \beta_{\tau}\right) \Longrightarrow \begin{cases} N\left(0, \frac{\tau(1-\tau)}{f_{u_{0\tau}}(0)^2}\begin{bmatrix} 1 & 0 \\ 0 & \Omega_{xx}^{-1}\end{bmatrix}\right) & \textit{for (I0)}, \\ N\left(0, \frac{\tau(1-\tau)}{f_{u_{0\tau}}(0)^2}\begin{bmatrix} 1 & 0 \\ 0 & V_{xx}^{-1}\end{bmatrix}\right) & \textit{for (MI)}, \\ f_{u_{0\tau}}(0)^{-1}\begin{bmatrix} 1 & \int J_x^c(r)' \\ \int J_x^c(r) & \int J_x^c(r) J_x^c(r)'\end{bmatrix}^{-1}\begin{bmatrix} B_{\psi_{\tau}}(1) \\ \int J_x^c(r) dB_{\psi_{\tau}}\end{bmatrix} & \textit{for (I1)}, \\ MN\left[0, \frac{\tau(1-\tau)}{f_{u_{0\tau}}(0)^2}\begin{bmatrix} 1 & 0 \\ 0 & \tilde{V}_{xx}^{-1}\end{bmatrix}\right] & \textit{for (ME)}, \end{cases}$$

where $V_{xx} = \int_0^{\infty} e^{rC} \Omega_{xx} e^{rC} dr, \tilde{V}_{xx} = \int_0^{\infty} e^{-rC} Y_C Y_C' e^{-rC} dr$ *and* $Y_C \equiv N\left(0, \int_0^{\infty} e^{-rC} \Omega_{xx} e^{-rC} dr\right).$

2.3 Sources of Nonstandard Distortion and Correction Methods

Theorem 2.1 shows that the limit distribution in the (I1) case is nonstandard and nonpivotal. To see the source of nonstandard distortion clearly, I further analyze the limit distribution of the slope coefficient estimator. For simplicity, assume $K = 1$ and $u_{xt} \sim mds\left(0, \Sigma_{xx}\right)$, then it is straightforward to show that

$$n\left(\widehat{\beta}_{1,\tau}^{QR} - \beta_{1,\tau}\right) \sim f_{u_{0\tau}}\left(0\right)^{-1} \frac{\int \bar{J}_x^c dB_{\psi_\tau}}{\int \left(\bar{J}_x^c\right)^2},$$

where $\bar{J}_x^c = J_x^c(r) - \int_0^1 J_x^c(r)dr$ is the demeaned OU process. Using the orthogonal decomposition of Brownian motion (Phillips, 1989) $dB_{\psi_\tau} = dB_{\psi_\tau . x} + \Sigma_{\psi_\tau x}\Sigma_{xx}^{-1}dB_x$ we have

$$n\left(\widehat{\beta}_{1,\tau}^{QR} - \beta_{1,\tau}\right) \sim f_{u_{0\tau}}\left(0\right)^{-1} \left[\frac{\int \bar{J}_x^c dB_{\psi_\tau . x}}{\int \left(\bar{J}_x^c\right)^2} + \left(\frac{\Sigma_{\psi_\tau x}}{\Sigma_{xx}}\right)\frac{\int \bar{J}_x^c dB_x}{\int \left(\bar{J}_x^c\right)^2}\right].$$

Note that:

$$\frac{\int \bar{J}_x^c dB_{\psi_\tau . x}}{\int \left(\bar{J}_x^c\right)^2} \equiv MN\left(0, \Sigma_{\psi_\tau . x}\left[\int \left(\bar{J}_x^c\right)^2\right]^{-1}\right),$$

with $\Sigma_{\psi_\tau . x} = Var\left(B_{\psi_\tau . x}\right) = \tau(1-\tau) - \Sigma_{xx}^{-1}\Sigma_{\psi_\tau x}^2$ and $\Sigma_{\psi_\tau x} = Cov\left(\psi_\tau\left(u_{0t\tau}\right), u_{xt}\right)$. Now assume a researcher uses the ordinary QR standard error $s.e(\widehat{\beta}_{1,\tau}^{QR}) = \tau(1-\tau)\hat{f}_{u_{0\tau}}\left(0\right)^{-1}\{\sum_{t=1}^n (x_{t-1}^{\mu})^2\}^{-1/2}$, where $x_{t-1}^{\mu} = x_{t-1} - T^{-1}\sum x_{t-1}$. Then with the standardized notation $\left(\bar{I}_x^c, W_x\right) = \Sigma_{xx}^{-1/2}\left(\bar{J}_x^c, B_x\right)$, the t-ratio becomes:

$$\begin{aligned} t_{\widehat{\beta}_{1,\tau}} = \frac{\left(\widehat{\beta}_{1,\tau}^{QR} - \beta_{1,\tau}\right)}{s.e(\widehat{\beta}_{1,\tau}^{QR})} &\sim \left[1 - \frac{\Sigma_{\psi_\tau x}^2}{\Sigma_{xx}\tau(1-\tau)}\right]^{1/2} N(0,1) + \frac{\Sigma_{\psi_\tau x}}{\left(\Sigma_{xx}\tau(1-\tau)\right)^{1/2}}\frac{\int \bar{I}_x^c dW_x}{\left[\int \left(\bar{I}_x^c\right)^2\right]^{1/2}} \\ &\sim \underbrace{\left[1 - \lambda\left(\tau\right)^2\right]^{1/2} Z}_{\text{standard inference}} + \underbrace{\lambda\left(\tau\right)\eta_{LUR}(c)}_{\text{nonstandard distortion}} \end{aligned} \tag{2.10}$$

where Z and $\eta_{LUR}(c)$ stand for a standard normal distribution and the local unit root t-statistics, respectively, and

$$\lambda\left(\tau\right) = -corr\left(1\left(u_{0t\tau} < 0\right)u_{xt}\right) \neq corr(u_{0t}, u_{xt}) := \phi.$$

Remark 2.2 *As the analytical expression (2.10) shows, the nonstandard distortion becomes greater with (i) smaller $|c|$ and (ii) larger $|\lambda\left(\tau\right)|$. Condition (i) is well known from the mean predictive regression literature where the distortion from the highly left-skewed feature of $\eta_{LUR}(c)$ with small $|c|$ has been studied. Condition (ii) is a special feature of nonstationary QR, see Xiao (2009) for strict unit root regressors and Maynard et al. (2011) for a local-to-unity regressor with an explanation of the nonstandard feature from this distortion. Note that*

$$\lambda\left(\tau\right) = \frac{-E\left[1\left(u_{0t\tau} < 0\right)u_{xt}\right]}{\{\Sigma_{xx}\tau(1-\tau)\}^{1/2}} = \frac{-E\left[1\left(u_{0t} < F_{u_0}^{-1}(\tau)\right)u_{xt}\right]}{\{\Sigma_{xx}\tau(1-\tau)\}^{1/2}}, \tag{2.11}$$

so the explicit source of distortion from persistence and nonlinear dependence is provided by this analysis. For exposition, assuming multivariate normal or t distributions for (u_{0t}, u_{xt}) *gives (see 1.1 of Lee; 2014),*

$$\lambda(\tau) = \frac{\phi\left(-\int_{-\infty}^{F_{u_0}^{-1}(\tau)} y f_{u_0}(y)\, dy\right)}{\sqrt{\tau(1-\tau)\Sigma_{xx}}} = \phi\frac{\tau}{\sqrt{\tau(1-\tau)\Sigma_{xx}}}\left(-E\left[u_0 | u_0 < F_{u_0}^{-1}(\tau)\right]\right) \tag{2.12}$$

where we clearly see the QR endogeneity $\lambda(\tau)$ *is a composite effect of linear dependence* (ϕ) *and the truncated mean of regression errors* ($-\int_{-\infty}^{F_{u_0}^{-1}(\tau)} y f_{u_0}(y)\, dy$). *Another interesting property is that* $|\lambda(\tau)| < |\phi|$ *can be shown for certain cases. This result indicates a inherent robustness of QR under persistence, i.e., less distortion than mean regression given an identical degree of persistence (c). In Table A.1 in Lee (2014), various* $\lambda(\tau)'s$ *are calculated when* $\phi = -0.95$, *and the relation* $|\lambda(\tau)| < |\phi|$ *figures. Moreover, the magnitude of* $|\lambda(\tau)|$ *gets larger (more distortion) as the tail gets heavier at the 5% quantile. Meanwhile,* $|\lambda(\tau)|$ *gets smaller for thicker tails at the median.*

Remark 2.3 *The commonly used lower tail dependence measure is* $\lambda_L = \lim_{\tau \to 0+} \lambda_L(\tau)$, *where*

$$\lambda_L(\tau) = P\left[u_{0t} < F_{u_0}^{-1}(\tau) | u_{xt} < F_{u_x}^{-1}(\tau)\right] = \tau^{-1} E\left[1\left(u_{0t} < F_{u_0}^{-1}(\tau)\right) 1\left(u_{xt} < F_{u_x}^{-1}(\tau)\right)\right].$$

Thus, the dependence measure $\lambda(\tau)$ *in (2.11) is different from both linear dependence* (ϕ) *and tail or quantile dependence* (λ_L *or* $\lambda_L(\tau)$). *The linear dependence* (ϕ) *affects the distortion of nonstationary mean regression, while* $\lambda(\tau)$ *contributes to the distortion in nonstationary QR. If we consider a quantile-quantile predictability (Han et al., 2014) or a extreme quantile version of it (Davis and Mikosch, 2009),* $\lambda_L(\tau)$ *or* λ_L *will play the contributing role for the distortion. The quantile-quantile predictability under the presence of persistent predictors will be an interesting topic for future research, wherein we would need to define a proper version of quantile for nonstationary processes.*

To correct the nonstandard distortion in (2.10), we may consider two approaches. The first is to construct a confidence interval (CI) for c, such as Stock's CI (1991), and correct the distortion through an induced CI for $\beta_{1,\tau}$. This type of Bonferroni methods are frequently used in predictive mean regression (e.g., Cavanagh et al.,1995; Campbell and Yogo, 2006). For a single local to unity (I1) predictor, Campbell and Yogo (2006) successfully correct the distortion, but lose their validity when predictor persistence belongs to (MI) or (I0) spaces. Cavanagh et al. (1995) still provides conservative size control but may become overly conservative for (MI) to (I0) regressors - see Phillips (2014). However, the Bonferroni methods based on a uniformly valid CI for c (e.g., Mikusheva, 2007), rather than Stock's CI, will provide validity over (MI) or (I0) spaces. I provide the simulation comparison of IVX-QR to this modified Bonferroni correction as well as the original Campbell-Yogo method (Figure 3). For multivariate nonstationary predictors (multiple c_i's), Bonferroni methods are somewhat difficult to use. In predictive QR, Maynard et al. (2011) employed the Bonferroni correction idea. The second approach to correct for nonstandard distortion, which this paper follows, is to use the IVX filtering technique (Magdalinos and Phillips, 2009). Methods based on the IVX filtering technique are discussed in the next section.

3 IVX-QR Methods

It is convenient to transform the model (2.6) to remove the intercept term:

$$y_{t\tau} = \beta'_{1,\tau} x_{t-1} + u_{0t\tau} \tag{3.1}$$

where $y_{t\tau} := y_t - \hat{\beta}^{QR}_{0,\tau}(\tau) = y_t - \beta_{0,\tau} + O_p(n^{-1/2})$ is the zero-intercept QR dependent variable. This is analogous to the demeaning process in the predictive mean regression in preparation for tests of the slope coefficient. Section 1.4 of Lee (2014) explains the validity of the dequantiling procedure and a possible inference on $\beta_{0,\tau}$.

3.1 IVX Filtering

This Section reviews a new filtering method, IVX filtering (Magdalinos and Phillips, 2009). The idea can be explained by comparing it to commonly used filtering methods. For simplicity, first assume x_t belongs to (I1). Filtering persistent data x_t to generate $\tilde{z}_t$ can be described as

$$\tilde{z}_t = F\tilde{z}_{t-1} + \triangle x_t$$

with a filtering coefficient F and first difference operator $\triangle$.

When $F = 0_K$ then $\tilde{z}_t = \triangle x_t$ and we simply take the first difference to remove the persistence in x_t. First differencing is the most common technique employed by applied researchers, and it leads to the (I0) limit theory in Theorem 2.1. Thus, the standard normal (or chi square) inference is achieved. The drawback to first differencing is the substantial loss of statistical power in detecting predictability of x_{t-1} on y_t. Taking the first difference of a regression equation makes both x_{t-1} and y_t much noisier and finding the relationship between two noisy processes is statistically challenging. In terms of convergence rate, the first difference reduces the n-rate (for the (I1) case) to the $n^{1/2}$-rate (for the (I0) case), thereby seriously diminishing local power. At the cost of this substantial loss, the first difference technique corrects the nonstandard distortion in (2.10).

When $F = I_K$ then $\tilde{z}_t = x_t$ so we use level data without any filtering. The statistical power is preserved in this way, since it is easy to detect if a persistent x_{t-1} has non-negligible explanatory power on noisy y_t. This is clear from the n-rate of convergence of (I1) limit theory (maximum rate efficiency) in Theorem 2.1. However, inference suffers from the size distortion in (2.10).

The main idea of IVX filtering is to filter x_t to generate $\tilde{z}_t$ with (MI) persistence - intermediate between first differencing and the use of levels data. In particular, we choose $F = R_{nz}$ as follows:

$$\tilde{z}_t = R_{nz}\tilde{z}_{t-1} + \triangle x_t,\ R_{nz} = I_K + \frac{C_z}{n^{\delta}}, \tag{3.2}$$
$$\text{where } \delta \in (0,1),\ C_z = c_z I_K,\ c_z < 0 \text{ and } \tilde{z}_0 = 0.$$

The parameters $\delta \in (0,1)$ and $c_z < 0$ are specified by the researcher. One practical suggestion is given in Section 4.1. As is clear from the construction, R_{nz} is between 0_K and I_K but closer to I_K especially for large n. This construction is designed to preserve local power as much as possible while achieving the desirable size correction. The $\tilde{z}_t$ essentially belongs to an (MI) process so the limit theory of the (MI) case in Theorem 2.1 is obtained by using $\tilde{z}_t$ as instruments. The IVX filtering exploits advantages both from using level (power) and the first difference (size correction) of persistent data. It leads to the intermediate signal strength $n^{(1+\delta)/2}$. At the cost of the slight reduction in convergence rate compared to the level data, the filtering achieves the desired size correction. The simulation in Section 4 shows that this cost may not be substantial. To summarize

Comparisons of level, first differenced and IVX-filtered data:

	level	first difference	IVX filtering
Discriminatory power	Yes	No	Yes
Size correction	No	Yes	Yes
Rate of convergence	n	$n^{1/2}$	$n^{(1+\delta)/2}$

Assume now that x_t falls into one of three specifications: (I0), (MI) and (I1). When x_t belongs to (I1), the IVX filtering reduces the persistence to (MI) as described above. If x_t belongs to (MI) or (I0), the filtering maintains the original persistence. This is how we achieve uniform validity over the range of (I0)-(I1). This automatic adjustment applies to several persistent predictors simultaneously, thereby accommodating multivariate persistent regressors. When x_t belongs to (ME), the IVX estimation becomes equivalent to OLS for the mean regression case (Phillips and Lee, 2014). The same principle works for QR, delivering uniformly valid inference in QR over (I0)-(ME) predictors (Proposition 3.1 and 3.2 below).

3.2 IVX-QR Estimation and Limit Theory

I propose new IVX-QR methods that are based on the use of IVX filtered instruments. Since the rate of convergence of IVX-QR will differ according to predictor persistence, I unify notation for the data with the following embedded normalizations:

$$\tilde{Z}_{t-1,n} := \tilde{D}_n^{-1}\tilde{z}_{t-1} \text{ and } X_{t-1,n} := \tilde{D}_n^{-1}x_{t-1}, \tag{3.3}$$

where, using notation $\alpha \wedge \delta = \min(\alpha, \delta)$,

$$\tilde{D}_n = \begin{cases} \sqrt{n} I_K & \text{for (I0)}, \\ n^{\frac{1+(\alpha\wedge\delta)}{2}} I_K & \text{for (MI) and (I1)}, \\ n^{(\alpha\wedge\delta)} R_n^n & \text{for (ME)}. \end{cases}$$

The unified normalizing matrix becomes one of these three specifications according to predictor persistence and the relation between α and δ (through $\alpha \wedge \delta$). This notation is convenient for presenting the IVX-QR limit theory (Theorem 3.1 and 3.2 below) but depends on the unknown localizing coefficient matrix C and the unknown rate parameter α. Thus Theorem 3.1 and 3.2 are not directly applicable for practical work. However, self normalized versions of the statistics (Proposition 3.1 and 3.2) have a chi-square limit theory free of these unknown parameters, providing a basis for the actual inference.

I also unify the different asymptotic moment matrices for the (MI) and (I1) cases:

$$V_{cxz} := \begin{cases} V_{zz}^x = \int_0^\infty e^{rC_z} \Omega_{xx} e^{rC_z} dr, & \text{when } \delta \in (0, \alpha \wedge 1), \\ V_{xx} = \int_0^\infty e^{rC} \Omega_{xx} e^{rC} dr, & \text{when } \alpha \in (0, \delta). \end{cases} \tag{3.4}$$

and

$$\Psi_{cxz} := \begin{cases} -C_z^{-1} \left\{ \Omega_{xx} + \int dJ_x^c J_x^{c\prime} \right\}, & \text{if } \alpha = 1, \\ -C_z^{-1} \left\{ \Omega_{xx} + C V_{xx} \right\}, & \text{if } \alpha \in (\delta, 1), \\ V_{cxz} = V_{xx} & \text{if } \alpha \in (0, \delta). \end{cases} \tag{3.5}$$

From the conditional moment restriction $E\left[\tau - 1\left(y_{t\tau} \leq \beta_{1,\tau}{}' x_{t-1}\right) | \mathcal{F}_{t-1}\right] = 0$, a natural procedure of estimating $\beta_{1,\tau}$ using IVX filtering is to minimize the L_2-distance of the sum of the empirical moment conditions that use IVX $\tilde{z}_{t-1}$ from information set $\mathcal{F}_{t-1}$.

Definition 3.1 (IVX-QR estimation) *The IVX-QR estimator $\hat{\beta}_{1,\tau}$ for $\beta_{1,\tau}$ is defined as*

$$\hat{\beta}_{1,\tau}^{IVXQR} = \arg\inf_{\beta_1} \frac{1}{2} \left(\sum_{t=1}^n m_t(\beta_1) \right)' \left(\sum_{t=1}^n m_t(\beta_1) \right), \tag{3.6}$$

where $m_t(\beta_1) = \tilde{z}_{t-1}(\tau - 1(y_{t\tau} \leq \beta_1{}' x_{t-1})) = \tilde{z}_{t-1} \psi_\tau(u_{0t\tau}(\beta_1))$.

The minimization (3.6) leads to the following approximate FOC:

$$\sum_{t=1}^n \tilde{Z}_{t-1,n} \left(\tau - 1\left(y_{t\tau} \leq \left(\hat{\beta}_{1,\tau}^{IVXQR} \right)' x_{t-1} \right) \right) = o_p(1). \tag{3.7}$$

The asymptotic theory of $\hat{\beta}_{1,\tau}^{IVXQR}$ follows from this condition. The next theorem gives the limit theory of the IVX-QR estimator under various degrees of predictor persistence.

Theorem 3.1 (IVX-QR limit theory)

$$\tilde{D}_n\left(\hat{\beta}_{1,\tau}^{IVXQR}-\beta_{1,\tau}\right) \Longrightarrow \begin{cases} N\left(0,\frac{\tau(1-\tau)}{f_{u_{0\tau}}(0)^2}\Omega_{xx}^{-1}\right) & \text{for (I0)}, \\ MN\left(0,\frac{\tau(1-\tau)}{f_{u_{0\tau}}(0)^2}\Psi_{cxz}^{-1}V_{cxz}\left(\Psi_{cxz}^{-1}\right)'\right) & \text{for (MI) and (I1)}, \\ MN\left(0,\frac{\tau(1-\tau)}{f_{u_{0\tau}}(0)^2}\left(\tilde{V}_{xx}\right)^{-1}\right) & \text{for (ME)}. \end{cases}$$

Unlike Theorem 2.1, the limit theory is (mixed) normal for all cases, and the limit variances are easily estimated. The self-normalized estimator given in the following theorem provides a convenient tool for unified inference across the (I0), (MI), (I1) and (ME) cases.

Proposition 3.1 (Self-normalized IVX-QR) *For (I0), (MI), (I1) and (ME) predictors,*

$$\widehat{f_{u_{0\tau}}(0)}^2(\tau(1-\tau))^{-1}(\hat{\beta}_{1,\tau}^{IVXQR}-\beta_{1,\tau})'\left(X'P_{\tilde{Z}}X\right)(\hat{\beta}_{1,\tau}^{IVXQR}-\beta_{1,\tau}) \Longrightarrow \chi^2(K),$$

where $X'P_{\tilde{Z}}X = X'\tilde{Z}\left(\tilde{Z}'\tilde{Z}\right)^{-1}\tilde{Z}'X = \left(\sum_{t=2}^n x_{t-1}\tilde{z}_{t-1}'\right)\left(\sum_{t=2}^n \tilde{z}_{t-1}\tilde{z}_{t-1}'\right)^{-1}\left(\sum_{t=2}^n x_{t-1}\tilde{z}_{t-1}'\right)'$ *and* $\widehat{f_{u_{0\tau}}(0)}$ *is any consistent estimator for* $f_{u_{0\tau}}(0)$ [3].

Using Proposition 3.1, we can test the linear hypothesis $H_0 : \beta_{1,\tau} = \beta_{1,\tau}^0$ for any given $\beta_{1,\tau}^0$. More generally, consider a set of r linear hypotheses $H_0 : H\beta_{1,\tau} = h_\tau$ with a known $r \times K$ matrix H and a known vector h_τ. In this case the null test statistics are formed as follows with the corresponding chi-square limit theory

$$\widehat{f_{u_{0\tau}}(0)}^2(\tau(1-\tau))^{-1}(H\hat{\beta}_{1,\tau}^{IVXQR}-h_\tau)'\{H\left(X'P_{\tilde{Z}}X\right)^{-1}H'\}^{-1}(H\hat{\beta}_{1,\tau}^{IVXQR}-h_\tau) \Longrightarrow \chi^2(r).$$

3.3 IVX-QR Inference: Testing Quantile Predictability

Theorem 3.1 and Proposition 3.1 allow for testing of a general linear hypothesis with multiple persistent predictors. The procedure (3.6) may be computationally demanding since the optimization of a nonconvex objective requires grid search with several local optima. Considering that the usual hypothesis of interest in predictive regression is the null of $H_0 : \beta_{1,\tau} = 0$, I propose an alternative testing procedure that is computationally attractive. Recall the DGP we impose is $y_{t\tau} = \beta_{1,\tau}'x_{t-1} + u_{0t\tau}$. Based on the fact that x_{t-1} and $\tilde{z}_{t-1}$ are "close" to each other, we use ordinary QR on $\tilde{z}_{t-1}$ to test $H_0 : \beta_{1,\tau} = 0$. Specifically, consider the simple QR regression procedure

$$\hat{\gamma}_{1,\tau}^{IVXQR} = \arg\min_{\gamma}\sum_{t=1}^n \rho_\tau\left(y_{t\tau}-\gamma_1'\tilde{z}_{t-1}\right).$$

We then have the following asymptotics of null test statistics:

[3] The kernel density estimation with standard normal kernel functions and Silverman's rule for the bandwidth choices are used in the simulation and empirical results below.

Theorem 3.2 *Under $H_0 : \beta_{1,\tau} = 0$,*

$$\tilde{D}_n \left(\hat{\gamma}_{1,\tau}^{IVXQR} - \beta_{1,\tau} \right) \Longrightarrow \begin{cases} N\left(0, \frac{\tau(1-\tau)}{f_{u_{0\tau}}(0)^2} \Omega_{xx}^{-1}\right) & \text{for (I0)}, \\ N\left(0, \frac{\tau(1-\tau)}{f_{u_{0\tau}}(0)^2} V_{cxz}^{-1}\right) & \text{for (MI) and (I1)}, \\ MN\left(0, \frac{\tau(1-\tau)}{f_{u_{0\tau}}(0)^2} \left(\tilde{V}_{xx}\right)^{-1}\right) & \text{for (ME)}. \end{cases}.$$

The above limit theory also holds under local alternatives of the form $H_0 : \beta_{1,\tau} = n^{-\nu} b_{1,\tau}$ with some $\nu > 0$. We achieve asymptotic normality of the null test statistics simply by replacing the regressor x_{t-1} with $\tilde{z}_{t-1}$. The final pivotal test statistics can be obtained by a similar self-normalization as given in the next theorem.

Proposition 3.2 *Under $H_0 : \beta_{1,\tau} = 0$,*

$$\widehat{f_{u_{0\tau}}(0)}^2 (\tau(1-\tau))^{-1} (\hat{\gamma}_{1,\tau}^{IVXQR} - \beta_{1,\tau})' \left(\tilde{Z}'\tilde{Z}\right) (\hat{\gamma}_{1,\tau}^{IVXQR} - \beta_{1,\tau}) \Longrightarrow \chi^2(K).$$

for (I0), (MI), (I1) and (ME) predictors.

Since QR algorithms are available in standard statistical software, Proposition 3.2 provides a uniform inference tool that involves easy computation. If we want to test the predictability of a specific subgroup among our predictors, say $H_0 : \beta_{11,\tau} = \beta_{12,\tau} = 0$, then the following test statistics with $H = \left[I_2, \mathbf{0}_{2\times(K-2)}\right]$ can be employed

$$\widehat{f_{u_{0\tau}}(0)}^2 (\tau(1-\tau))^{-1} (H\hat{\gamma}_{1,\tau}^{IVXQR})' \{H \left(\tilde{Z}'\tilde{Z}\right)^{-1} H'\}^{-1} (H\hat{\gamma}_{1,\tau}^{IVXQR}) \Longrightarrow \chi^2(2).$$

4 On the Choice of IVX-QR Filtering Parameters and Simulation

In this Section, I discuss the proper choice of parameters (C_z, δ) used in the IVX construction. Theorems 3.1 and 3.2 show that, for a given C_z, a larger IVX persistence (δ) leads to more local power while a smaller δ may achieve better size corrections. I suggest a practical choice rule and provide confirmatory simulation evidence obtained by IVX-QR tests based on this rule.

4.1 A Practical Rule

The idea uses the analytical formula (2.10) where the QR t-ratio is shown as $t_{\widehat{\beta}_{1,\tau}} \sim (1-\lambda(\tau)^2)^{1/2} Z + \lambda(\tau)\eta_{LUR}(c)$ with $\eta_{LUR}(c) = (\int (\bar{I}_x^c)^2)^{-1/2} \int \bar{I}_x^c dW_x$. The distributional properties of $\eta_{LUR}(c)$ are well understood. It is easy to simulate the distribution for a given c. The idea of IVX-QR filtering is to reduce the regressor persistence ($|c|$) in order to remove the nonstandard distortion arising from $\lambda(\tau)\eta_{LUR}(c)$, which is a composite effect of QR endogeneity and nonstationary distortion:

$$\underbrace{\lambda(\tau)}_{\text{QR endogeneity}} \times \underbrace{\eta_{LUR}(c)}_{\text{nonstationary distortion}}.$$

The essential AR(1) parameter of the IVX-filtered regressor (in a bivariate setting) is $R_{nz} = 1 + \frac{c_z}{n^\delta} = 1 + \frac{c(\delta,n)}{n}$ where $c(\delta,n) = n^{1-\delta}c_z$. Thus, for any $\delta \in (0,1)$, the induced local-to-unity parameter $c(\delta,n) \to -\infty$ as $n \to \infty$, thereby letting $\eta_{LUR}(c(\delta,n)) \Longrightarrow Z \sim N(0,1)$ (Phillips, 1987). The standard normal limit theory for $t_{\widehat{\beta}_{1,\tau}}$ will therefore hold for any $\delta \in (0,1)$ as $n \to \infty$. For a given finite n, the filtering parameter choice can be interpreted as showing how to choose a "proper minus infinity" $(n^{1-\delta}c_z)$ to deliver the standard normal limit theory of $t_{\widehat{\beta}_{1,\tau}}$. If δ is small enough, $c(\delta,n) = n^{1-\delta}c_z$ will be close enough to $-\infty$, however, smaller δ result in a loss of local power. For a given c_z and n, we want to increase $\delta < 1$ up a threshold value where the distortion from $\lambda(\tau)\,\eta_{LUR}(c)$ is still acceptable.

I suggest a practical rule to choose $\delta \in (0,1)$ with a normalized $c_z = -5$. In a local to unity regressor setting, it is possible to find a numerical value of $c(\delta,n) = c$ that controls the size distortion after imposing an acceptable Type I error bound. For example, Campbell and Yogo (2006) suggested 7.5%, for a nominal 5% level. More conservative bounds can be employed according to the purpose of the researchers. By simulating $t_{\widehat{\beta}_{1,\tau}}$ with various choices of $c(\delta,n)$ and $\lambda(\tau)$, we could plot the asymptotic size of nominal α t-test as a function of $c(\delta,n)$ and $\lambda(\tau)$:

$$AsySIZE\left(c(\delta,n), \lambda(\tau); \alpha\right) = \Pr\left(\left|(1-\lambda(\tau)^2)^{1/2} Z + \lambda(\tau)\,\eta_{LUR}(c)\right| > Z_{\alpha/2}\right).$$

For 5% level tests, I report the values of $c(\delta,n)$ and $\lambda(\tau)$ providing the desirable level of tail approximations, where the strict error bound is imposed to achieve the nominal size. They are tabulated in Table A.3 in the supplement (Lee, 2014). Using the table we could pick the proper $c(\delta,n)$ and equating $c(\delta,n) = n^{1-\delta}(c_z)$. The corresponding δ is therefore obtained by $\delta = 1 - (\log(-c(\delta,n)) - \log(5))/\log n$.

Among the two distortion components $\lambda(\tau)$ and $\eta_{LUR}(c)$, we only use data information contained in the estimated QR endogeneity $\hat{\lambda}(\tau)$. The estimation is straightforward by the regression residuals. The reliable performances are reported in Table A.2 (Section 1.1 of Lee; 2014). The choice based on $\hat{\lambda}(\tau)$ is thus quantile-dependent, and it allows us to pick the correct δ from data information. We do not use any estimation of c but rather directly impose an acceptable value of $c(\delta,n) = c$ to the IVX filtering mechanism. Not using any estimated c (which is not consistently estimable) is a key idea to avoid the invalidity issues with I(0)-MI predictors, raised in the recent literature (e.g., Mikusheva, 2007; Phillips, 2014).

In sum, given the data:

1. Obtain $\hat{\lambda}(\tau)$ by estimating $-corr\left(1\left(u_{0t\tau} < 0\right), u_{xt}\right)$ using the regression residuals.

2. Find a suggested $c(\delta,n)$ from Table A.3 and get $\delta = 1 - (\log(-c(\delta,n)) - \log(5))/\log n$.

3. Using the δ and $c_z = -5$, perform the IVX-QR estimation and tests.

In multiple predictor scenarios, we could use $\delta = \min(\delta_1, ..., \delta_K)$ to provide safe size control, where each δ_k is chosen by the above rule from the corresponding regressor x_k. The performances are investigated in the next Section.

The Matlab codes containing the built-in computation of $\hat{\lambda}(\tau)$ and the automatic IVX-QR correction based on the corresponding δ are used in the simulation and empirical applications below. The codes are available from the author's web page[4].

4.2 Simulation

I conduct simulations to examine the numerical performances of IVX-QR inference methods based on the practical rule suggested in Section 4.1. Using the common simulation designs, I confirm the validity of the suggested choice of (C_z, δ) and the reliable IVX-QR performances.

The following DGP is imposed:

$$\begin{aligned} y_t &= \beta_{0,\tau} + \beta'_{1,\tau} x_{t-1} + u_{0t,\tau}, \\ x_t &= \mu_x + R x_{t-1} + u_{xt}. \end{aligned} \tag{4.1}$$

where $\mu_x = 0$, $R = I_K + n^{-1}C$ and

$$u_t = (u_{0t}\ u'_{xt})' \sim iid\ F_u\left(0_{(K+1)\times 1}, \Sigma_{(K+1)\times(K+1)}\right). \tag{4.2}$$

The IVX is constructed as (3.2) using the practical rule of choice: δ is picked up from the look-up table A.3 based on $\hat{\lambda}(\tau)$,where C_z is normalized to $-5I_K$. The procedure is automatically built-in to the Matlab simulation codes. The tests use Proposition 3.2.

4.2.1 Median/Mean tests with a Single Persistent Predictor

I begin with a single ($K = 1$) persistent regressor. Although IVX-QR methods allow testing for predictability at various quantile levels, I focus on the median to compare its performance to that of existing mean predictability tests. In particular, the IVX-QR median test is compared to the methods of Campbell and Yogo (2006; CY-Q) and a modified version of CY-Q test (Modified CY-Q). The invalidity of the original CY-Q test for I(0)-MI regressors has been recently reported and a modification based on a uniformly valid CI (e.g., Mikusheva, 2007) was suggested in the literature (Phillips, 2014). I include both versions of CY-Q tests in the following simulation.

To investigate empirical size and power performances, I generate a sequence of local alternatives with $H_{\beta_{1n}} : \beta_{1n} = \frac{b}{n}$ in (4.1) for integer values $b \geq 0$ ($\tau = 0.5$ is suppressed) and observe the performances of the IVX-QR, CY-Q and modified CY-Q tests. Note that $b = 0$ case provides size performances ($H_0 : \beta_{1,\tau} = 0$), while $b > 0$ illustrates power results ($H_1 : \beta_{1,\tau} > 0$). Considering that much applied work uses the intercept term in the stock return regression (non-zero excess mean/median return), IVX-QR with dequantiling, as in (3.1), is compared to the CY-Q and modified CY-Q tests with demeaning. For the distribution F_u in (4.2), I employ normal and

[4] https://sites.google.com/site/jihyung412/research

t-distributions, with the correlation matrix

$$\Sigma = \begin{pmatrix} 1 & \phi \\ \phi & 1 \end{pmatrix}, \ \phi = -0.95.$$

This value of ϕ reflects the realistic error correlation in predictive regressions, such as dividend-price ratio, and has been frequently employed in the predictive regression literature.

Figures 1-3 illustrate the results. For I(1) ($R = 0.98$) predictors with normal errors, the CY-Q and modified CY-Q tests perform better than IVX-QR, which is expected. The test performance rankings with MI and I(0) ($R = 0.9$ and 0.84) with normal errors are mixed. For heavy-tailed scenarios (t(3)-t(1)), IVX-QR shows the best performance across all scenarios. Note that the invalidity of the original CY-Q for I(0) case (test size shrinking to zero) appears in Figure 3, again confirming the results in Phillips (2014) and Janson and Moreira (2006; Table I). The modified CY-Q shows the validity in all cases. The power loss of the modified CY-Q relative to the original CY-Q may be due to a particular Bonferroni correction employed here. The original CY-Q numerically found the combination of α_1 (size for R) and α_2 (size for β conditional on R) to maximize efficiency. The same efficiency maximization for the modified CY-Q is possible but is beyond the scope of this paper. We however expect the same ranking patterns will hold by comparing the best performances of two CY-Q's (CY-Q when $R = 0.98$ and 0.9, Modified CY-Q when $R = 0.84$) to IVX-QR's.

- FIGURES 1-3 HERE -

In summary, IVX-QR testing with a single persistent predictor is competitive, especially when we have heavy-tailed errors. All three tests perform well in terms of size and power except for the original CY-Q test in cases of stationary predictors. The IVX-QR test can easily employ multiple persistent predictors. In addition, the IVX-QR test can analyze the predictability of other quantiles in addition to the median, providing greater applicability for prediction tests. Size properties of IVX-QR prediction tests on various quantiles are analyzed in the next section.

4.2.2 Size Properties of Prediction Tests on Various Quantiles

Few studies have considered predicting other quantile levels of financial returns, such as the tail or shoulder (for exceptions, see Maynard et al., 2011; Cenesizoglu and Timmermann, 2008). This paper develops a valid method to test various quantile predictability of asset returns in the presence of multiple persistent predictors. In this Section, I focus on large sample performance ($n = 700$) to guarantee accurate density estimation at the tails, e.g., the 5% quantile. Imprecise density estimation at tail quantiles with finite sample size is a common problem in QR. Large sample sizes are often available in financial applications.

The simulation environment used to test the size properties of various quantile predictions is similar to that of the earlier Section. Dequantiling in (3.1) and the practical rule for (C_z, δ) based on the estimated $\lambda(\tau)$ are used for all IVX-QR simulations. The persistence parameter c_i is selected from $\{0, -2, -5, -7, -70\}$. This set represents a set of persistent predictors including $R = 0.9$

(MI) through $R = 1$ (unit root). Normal and t-distributions are used for F_u and the number of replications is 1000. All null test statistics use the same hypothesis: $H_0 : \beta_{1,\tau} = 0$ with a nominal size of 5%. The size performances exceeding 10% are shown in bold, which can be considered as severe size distortions.

I first investigate the size properties of ordinary QR methods. Table 1 below summarizes the size properties of ordinary QR t-statistics in (2.10) with a single persistent predictor when $\phi = -0.95$. The nonstandard distortion increases with more persistent predictors (smaller c). As Remark 2.2 suggests, the tail structure of F_u significantly affects the magnitude of the size distortion. For a t-distribution with heavier tails (smaller degrees of freedom), more severe size distortion arises at the tail than at the median, while the tendency does not impact normally distributed errors (thin tails). The overall results indicate the invalidity of the ordinary QR technique in the presence of persistent predictors, reassuring the findings of Xiao (2009) and Maynard et al. (2011).

The size performances of the IVX-QR methods are reported in Table 2. The size corrections are remarkable, confirming the validity of IVX-QR methods at various quantiles. A few tail cases with pure unit root regressors show mild over-rejections, but the distortions are substantially smaller than those of the ordinary QR. These mild over-rejections increase with heavier tails (t(3) and t(2) errors) as expected. The simulation results indicate that the IVX-QR correction methods with the choice rule of (C_z, δ) from Section 4.1 control test sizes well across most quantiles.

- TABLES 1 and 2 HERE -

I now consider the predictive QR scenario with multiple persistent predictors ($K = 2$). This scenario has rarely been explored but is relevant in empirical practice (e.g., book-to-market ratio and Treasury bill rate). To avoid lengthy documentation, I borrow a calibration technique for the innovation structure. In the empirical Section, specification with two predictors of book-to-market ratio and Treasury bill rate is shown to predict stock returns at various quantile levels. To support the empirical finding, the estimated correlation of the predictive QR application is used:

$$\Sigma = \begin{pmatrix} 1 & -0.78 & -0.17 \\ -0.78 & 1 & 0.21 \\ -0.17 & 0.21 & 1 \end{pmatrix}.$$

For bivariate predictor persistence, c_1 is set to 0 and c_2 is selected from $\{-2, -5, -7\}$.

Table 3 shows the size properties of ordinary QR test statistics. The size distortion is large when there are multiple persistent predictors, which corroborates the benefits of the IVX-QR method's validating inference under multiple manifestations of predictor persistence. Table 4 shows acceptable size results of the IVX-QR tests at various quantile levels. The size correction works well for most quantiles except a few tail cases. The results at inner quantiles from 0.2 to 0.8 are satisfactory. The IVX-QR corrections for multiple persistent predictors at $\tau = 0.05$ or 0.1 require further investigation. Even though there are some improved size controls over the ordinary QR (e.g., from 17.8% to 11.8% at $\tau = 0.1$ with t(3) errors), the tail case performances suggest a need for new methods to handle extremal quantiles under persistence. One potential solution could be the use of a recent development in extremal QR limit theory (e.g., Chernozhukov, 2005; Chernozhukov and Fernandez-Val, 2012). I leave this for future research.

- TABLES 3 and 4 HERE -

In summary, IVX-QR methods based on the practical rule of choice of (C_z, δ) demonstrate reliable size performances for most relevant specifications with single and multiple persistent predictors, except for a few extreme cases. More comprehensive simulation results are available from the online supplement (Lee, 2014). The practical benefits of IVX-QR inference will be illustrated through empirical examples in the next Section.

5 Quantile Predictability of Stock Returns

It is often standard practice to test stock return predictability using various economic and financial state variables as predictors. There is considerable disagreement in the empirical literature as to the predictability of stock returns when using a predictive mean regression framework (e.g., Campbell and Thompson, 2007; Goyal and Welch, 2007). In this section, I show empirical results of stock return *quantile* prediction tests using IVX-QR. Excess stock returns are measured by the difference between the S&P 500 index including dividends and the one month Treasury bill rate. I focus on eight persistent predictors: dividend price (*d/p*), earnings price (*e/p*), book to market (*b/m*) ratios, net equity expansion (*ntis*), dividend payout ratio (*d/e*), T-bill rate (*tbl*), default yield spread (*dfy*), term spread (*tms*) and various combinations of the above variables. The full sample period is January 1927 to December 2005. These data sets are standard and have been extensively used in the predictive regression literature. Cenesizoglu and Timmermann (2008) and Maynard et al., (2011) recently used the same data set in a QR framework[5]. Following Cenesizoglu and Timmermann (2008), I classify the predictors into three categories.

Valuation ratios	Bond yield measures	Corporate finance variables
dividend-price ratio (*d/p*)	three-month T-bill rate (*tbl*)	dividend-payout ratio (*d/e*)
earnings-price ratio (*e/p*)	term spread (*tms*)	net equity expansion (*ntis*)
book-to-market ratio (*b/m*)	default yield (*dfy*)	

I employ the IVX-QR methods to illustrate the benefits of these new methods. In particular, I first investigate the quantile predictability of stock returns using individual predictors and then analyze the improved predictive ability of possible combinations of predictors.

The null test statistics in Proposition 3.2 is used with the choice rule of filtering parameters (δ, C_z) from Section 4. Table 5 below reports the univariate regression results, where p-values (%)

[5] I thank Yini Wang for providing the data set. For detailed constructions and economic foundations of the data set, see Goyal and Welch (2007). Note that Maynard et al. (2011) and Cenesizoglu and Timmermann (2008) also considered stationary predictors other than the eight persistent predictors I use.

are rounded to one decimal place for exposition. The results shown in bold imply the rejection of the null hypothesis of no predictability at the 5% level.

- TABLE 5 HERE -

The result is roughly consistent with the results of Maynard et al. (2011) and Cenesizoglu and Timmermann (2008). I find significant lower quantile predictive ability for the *d/e* and middle quantile predictive power for the *tbl.* Evidence of both lower and upper quantile predictability from *b/m* and *dfy* are provided. Overall, I find little evidence of predictability at the median except *tbl.* The results confirm the weak predictability at the mean/median of stock returns, the stronger forecasting capability at quantiles away from the median and several quantile specific predictors.

For multivariate regression applications, I use selective predictor combinations for illustrative purposes. The selection scheme is as follows: First, I ignore the predictability evidence from the univariate regression results at the first and last two quantiles ($\tau = 0.05, 0.1, 0.9$ and 0.95), where the size control is not guaranteed for a few extreme cases from the simulation evidences. I choose significant predictors at other quantiles where at least two consecutive predictability evidences are detected (*d/e, b/m, tbl,* and *dfy* in this instance). Second, I classify the chosen predictors into three groups - Group L, Group M and Group B which are explained below.

Group L : lower quantile predictors	Group M : middle quantile predictors	Group B : lower & upper quantile predictors
d/e	*tbl*	*b/m and dfy*

Finally, I select one predictor from each group to produce a bivariate predictor and choose predictor combinations exhibiting little evidence of comovement between the predictors. Evidence of comovement between predictors does not completely reduce the appeal of the combinations; however, we may prefer less-comoving systems for better forecast models[6].

I employ two diagnostic tests to observe evidence of comovement between persistent predictors: (i) the correlation of x_{t-1}, and (ii) the cointegration tests between x_{t-1}. The two measures will provide evidence of comovement between all (I0)-(ME) predictors (see Section 3.2 of Lee; 2014). I find little evidence of comovement between (*bm, tbl*) and (*d/e, tbl*).

The above selection scheme is used primarily for illustrative purposes, and I do not rule out the possibility of significant results from other combinations[7]. However, it is partly justifiable. For example, both *dfy* and *tbl* are bond yield measures that likely co-move, while *tbl* is a macro variable that may have different patterns. If we choose between (*dfy, tbl*) and (*bm, tbl*), the above rationale recommends (*bm, tbl*) because they share fewer common characteristics. Diagnostic tests indicate evidence of larger comovement between (*dfy, tbl*) than that between (*bm, tbl*). Therefore I focus on two combinations; (*bm, tbl*) and (*d/e, tbl*).

[6] Phillips (1995) provided robust inference methods in cointegrating mean regression models with possibly comoving persistent regressors (FM-VAR regressions). Introducing the robust feature into the current framework (allowing singular Ω_{xx} in (2.7)) will be left for future research.

[7] Results for other combinations are readily available upon request.

From Table 6 below, I confirm that the two combinations, (*bm, tbl*) and (*d/e, tbl*) are jointly significant at various quantiles with stronger evidence than that of univariate regressions. Many existing studies only considered a single persistent predictor. The results below illustrate the possibility of better forecast models with multiple persistent predictors that are not subject to spurious forecasts. We can proceed with more than two predictor models in a similar way.

-TABLE 6 HERE-

I run the prediction tests to significant predictors from Table 6 for post-1952 data. Many papers have reported that the stock return predictability becomes much weaker from January 1952 to December 2005 (see Campbell and Yogo, 2006; Kostakis et al. 2012). Papers have often argued that the disappearance of predictability was likely due to structural changes or improved market efficiency. Table 7 below shows weaker predictability evidence, but some differences to mean predictive regressions still exist. For example, Campbell and Yogo (2006) reported the predictive ability of the *tbl* during this sub-period, while Kostakis et al., (2012) concluded that the predictability from the variable disappears. I find significant results from *tbl* at lower to middle quantiles while little evidence at upper quantiles.

-TABLE 7 HERE-

I proceed to the tests with two predictors to confirm the earlier results on (*bm, tbl*) and (*d/e, tbl*). From Table 8 below, we see empirical support for stock return forecast models for post-1952 periods, using (*bm, tbl*) and (*d/e, tbl*). It turns out that the combination of one valuation ratio (*b/m*) and a macro variable (*tbl*), or the latter with a corporate finance variable (*d/e*), provide a potentially improved forecast model for stock returns. IVX-QR corrections ensure that the predictability results are not spurious.

-TABLE 8 HERE-

To summarize the empirical findings, I show that commonly used persistent predictors have greater predictive capability at some specific quantiles of stock returns, where the predictability from a given predictor tends to locate at lower or upper quantiles of stock returns but mostly disappears at the median. A partial answer to the empirical puzzle of stock return mean/median predictability may be provided. The significant predictors for specific quantiles of stock returns can play important roles in risk management and portfolio decision applications. I also find that by employing some combination of persistent variables as predictors, forecasting capability at most quantiles can be substantially enhanced relative to a model with a single predictor. The predictive performance of a specific combination, such as T-bill rate (*tbl*) and book-to-market ratio (*b/m*), remains high even during the post-1952 period. The improved in-sample quantile forecast results are not spurious because the IVX-QR methods control the size distortion. This finding is new in the literature, suggesting the potential for improved stock return forecast models.

6 Conclusion

This paper develops a new theory of inference for quantile regression (QR). I propose methods of robust inference which involve the use of QR with filtered instruments that lead to a new procedure called IVX-QR. These new methods accommodate multiple persistent predictors and they have uniform validity under various degrees of persistence. Both properties offer great advantages for empirical research in predictive regression.

In the empirical application of these methods, the tests confirm that commonly used persistent predictors have significant in-sample forecasting capability at specific quantiles, mostly away from the median. The IVX-QR methods allow the investigator to cope with quantile specific predictability of stock returns without exposing the outcomes to spurious effects from multiple persistent predictor. The enhanced predictive ability from combinations of persistent predictors suggests there is scope for further improvement in time series forecasting applications.

Several directions of future research are of interest. One is out-of-sample forecasting based on the IVX-QR methods. Explicit use of IVX-QR forecasts in portfolio decision making and risk analysis can be also studied. IVX-QR inference at extreme quantiles requires further theoretical investigations.

Table 1: Size Performances (%) of Ordinary QR ($n = 700$, $S = 1000$)

Normally distributed errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c = 0$	**14.5**	**13.5**	**15.0**	**15.5**	**16.3**	**17.8**	**17.0**	**16.3**	**16.0**	**13.1**	**14.5**
$c = -2$	**10.6**	**10.0**	**11.0**	**12.5**	**12.9**	**11.9**	**11.9**	**11.4**	9.8	9.6	**11.5**
$c = -5$	8.7	8.1	8.3	8.4	7.7	9.4	8.9	9.8	7.5	**11.0**	**10.6**
$c = -7$	9.7	9.2	7.4	6.3	7.4	6.2	6.2	7.1	7.4	7.8	9.5
$c = -70$	6.7	6.1	6.0	4.2	5.2	3.8	4.2	4.6	4.6	5.6	8.2

t(3) errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c = 0$	**18.1**	**17.0**	**13.5**	**12.5**	**12.3**	**11.5**	**11.9**	**12.9**	**14.4**	**17.1**	**17.6**
$c = -2$	**14.4**	**13.3**	**11.1**	8.5	8.5	9.1	9.2	**11.2**	**11.3**	**13.3**	**15.0**
$c = -5$	**11.8**	**11.5**	8.1	9.1	6.6	8.1	7.4	6.5	8.4	**11.4**	**12.9**
$c = -7$	**12.5**	8.1	**10.1**	7.5	5.5	7.1	5.8	6.9	6.4	9.4	**13.5**
$c = -70$	**10.0**	8.9	5.4	4.8	5.5	3.6	4.2	4.7	6.5	7.8	9.3

t(2) errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c = 0$	**20.3**	**16.1**	**11.3**	**11.1**	9.1	8.5	8.5	**10.4**	**12.3**	**14.2**	**16.2**
$c = -2$	**15.8**	**12.4**	**10.7**	8.3	8.6	5.8	6.7	8.1	**11.8**	**12.0**	**15.7**
$c = -5$	**13.0**	9.0	6.7	6.2	5.4	5.4	4.7	6.5	8.4	**12.9**	**13.7**
$c = -7$	**12.2**	**10.7**	7.8	5.6	4.1	3.9	5.8	5.3	7.9	**10.7**	**13.9**
$c = -70$	7.2	7.1	6.8	4.5	4.6	3.4	3.7	5.3	7.0	8.3	8.6

Table 2: Size Performances (%) of IVX-QR ($n = 700$, $S = 1000$)

Normally distributed errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c = 0$	8.3	8.5	9.4	8.7	6.0	6.6	8.6	9.9	8.8	7.9	**10.3**
$c = -2$	7.9	6.1	5.0	5.5	4.6	4.1	4.5	5.3	5.8	6.0	7.7
$c = -5$	6.4	5.4	4.6	4.9	4.8	2.7	4.7	5.3	4.7	5.4	7.0
$c = -7$	6.6	7.1	4.7	5.9	4.3	3.9	4.0	4.5	4.6	4.8	6.2
$c = -70$	5.1	5.7	4.3	3.5	3.4	3.0	3.4	5.1	4.6	5.1	7.2

t(3) errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c = 0$	**11.1**	**12.8**	9.1	7.8	7.5	6.8	6.1	7.7	7.6	**10.1**	**11.8**
$c = -2$	9.3	8.1	6.3	5.2	4.9	4.1	5.4	5.8	5.5	7.8	8.0
$c = -5$	5.8	7.5	5.9	5.2	4.0	3.7	4.2	3.9	5.9	7.4	7.6
$c = -7$	8.5	6.5	5.0	4.5	3.9	4.3	3.9	4.3	5.3	5.5	7.8
$c = -70$	7.1	8.0	5.7	3.6	4.2	3.4	4.6	5.0	5.2	5.9	7.0

t(2) errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c = 0$	**13.9**	**10.8**	8.0	7.8	6.1	5.7	7.5	8.9	9.9	7.8	**10.3**
$c = -2$	8.8	7.2	7.1	4.6	4.8	4.8	3.9	4.7	8.6	8.6	6.4
$c = -5$	7.8	6.6	6.8	5.3	4.8	4.2	4.9	5.0	7.4	7.7	7.5
$c = -7$	8.0	7.4	6.3	4.6	3.3	3.8	4.1	4.2	7.1	7.7	8.6
$c = -70$	5.1	7.1	6.2	4.4	4.0	3.6	3.3	5.2	7.0	6.6	6.7

Table 3: Size Performances (%) of Ordinary QR ($n = 700$, $S = 1000$)

Normally distributed errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c_1 = 0$, $c_2 = -2$	**13.9**	**11.8**	**12.8**	**13.0**	**12.6**	**12.6**	**13.0**	**10.7**	**12.4**	**11.1**	**13.7**
$c_1 = 0$, $c_2 = -5$	**14.0**	**12.2**	**12.3**	**12.0**	9.8	**10.4**	**13.1**	**14.6**	**11.0**	**12.7**	**12.3**
$c = 0$, $c_2 = -7$	**13.2**	**13.4**	9.8	9.7	**13.0**	**11.1**	**12.9**	**11.0**	**10.1**	**12.6**	**12.6**

t(3) errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c_1 = 0$, $c_2 = -2$	**20.4**	**17.8**	**13.6**	8.9	8.9	8.7	**10.0**	**10.6**	**11.8**	**13.5**	**19.8**
$c_1 = 0$, $c_2 = -5$	**23.9**	**14.6**	**11.5**	**10.1**	8.2	8.6	9.1	**10.1**	**12.6**	**19.6**	**17.5**
$c = 0$, $c_2 = -7$	**18.4**	**15.9**	**10.6**	9.2	8.3	8.4	7.4	**11.4**	**10.9**	**14.4**	**19.3**

Table 4: Size Performances (%) of IVX-QR ($n = 700$, $S = 1000$)

Normally distributed errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c_1 = 0$, $c_2 = -2$	**12.9**	**11.0**	9.9	9.6	8.1	7.6	7.8	8.0	9.7	9.0	**12.6**
$c_1 = 0$, $c_2 = -5$	**12.9**	**10.1**	8.4	7.8	7.1	6.1	7.9	7.8	7.8	**10.4**	**11.4**
$c = 0$, $c_2 = -7$	**13.3**	**11.3**	9.1	7.7	7.0	6.6	8.3	9.4	7.8	**10.2**	**10.4**

t(3) errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c_1 = 0$, $c_2 = -2$	**18.7**	**11.8**	8.8	9.4	6.4	4.4	6.3	8.5	9.1	**11.9**	**15.9**
$c_1 = 0$, $c_2 = -5$	**18.6**	**11.7**	9.7	5.6	7.8	4.5	6.6	8.2	9.0	**10.6**	**14.4**
$c = 0$, $c_2 = -7$	**14.7**	**14.1**	9.6	8.0	5.7	3.7	6.2	9.5	9.8	**10.1**	**17.4**

Table 5: p-values(%) of quantile prediction tests (1927:01-2005:12)

Univariate regressions with each of the eight predictors: *d/p, d/e, b/m, tbl, dfy, ntis, e/p, tms*

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	**0.3***	**0.0***	**0.2***	12.6	75.1	61.1	15.9	**0.4***	6.4	5.0	**0.0***
d/e	**0.0***	**0.0***	**0.0***	**0.0***	**1.6***	38.4	53.6	9.9	**2.9***	**0.0***	**0.0***
b/m	**0.1***	**0.0***	**0.1***	**0.5***	51.5	79.7	12.8	**0.1***	**2.6***	**0.7***	**0.8***
tbl	59.8	68.6	5.7	**0.7***	**0.4***	**1.2***	**0.7***	9.8	**1.8***	**0.1***	14.0
dfy	**0.0***	**0.0***	**0.0***	**0.0***	**3.2***	75.2	**2.6***	**0.0***	**0.0***	**0.0***	**0.0***
e/p	91.7	90.2	57.4	76.1	80.3	82.4	97.6	53.7	**0.4***	27.9	25.4
ntis	5.3	**0.3***	**0.3***	27.2	19.0	82.9	74.2	58.5	83.7	93.3	96.6
tms	16.8	5.8	22.1	61.3	23.6	36.1	24.4	70.6	17.7	**0.0***	**0.0***

Table 6: p-values(%) of quantile prediction tests (1927:01-2005:12)

Multivariate regressions with two predictors: (*b/m, tbl*) *and* (*d/e, tbl*)

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
b/m, tbl	**0.1***	**0.0***	**0.1***	**2.3***	19.5	**4.7***	**1.5***	**0.0***	**0.1***	**0.0***	**0.8***
d/e, tbl	**0.0***	**0.0***	**0.0***	**0.0***	**0.0***	5.5	**3.8***	14.1	**0.0***	**0.0***	**0.0***

Table 7: p-values(%) of quantile prediction tests (1952:01-2005:12)

Univariate regressions with each of the eight predictors: *d/p, d/e, b/m, tbl, dfy*

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	73.5	71.6	5.2	29.6	8.0	84.6	**1.4***	**0.1***	16.6	75.1	16.1
d/e	9.2	**2.5***	**4.8***	**2.0***	**0.1***	**2.4***	80.2	6.6	10.8	**0.2***	85.8
b/m	**0.1***	97.6	8.7	34.3	10.3	53.5	69.2	12.2	32.0	20.6	19.0
tbl	6.7	**0.3***	**0.0***	**0.0***	**0.1***	**0.2***	**0.7***	20.0	23.6	8.7	19.3
dfy	24.7	88.0	58.4	45.4	58.3	6.9	**2.3***	**1.8***	**0.5***	**0.2***	**0.1***

Table 8: p-values(%) of quantile prediction tests (1952:01-2005:12), $\delta = 0.5$

Multivariate regressions with two predictors: $(b/m,\ tbl)$ *and* $(d/e,\ tbl)$

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$b/m,\ tbl$	**0.0***	**1.3***	**3.5***	**0.0***	**0.6***	**2.9***	**0.7***	**1.4***	**0.7***	**0.1***	6.1
$d/e,\ tbl$	20.5	**1.1***	**0.0***	**0.0***	**0.0***	**0.0***	**1.0***	**1.0***	39.4	**0.0***	6.6

Figure 1: $c = -5$ $(n = 250,\ R = 0.98)$ with Normal, t(3), t(2) and t(1) innovations.

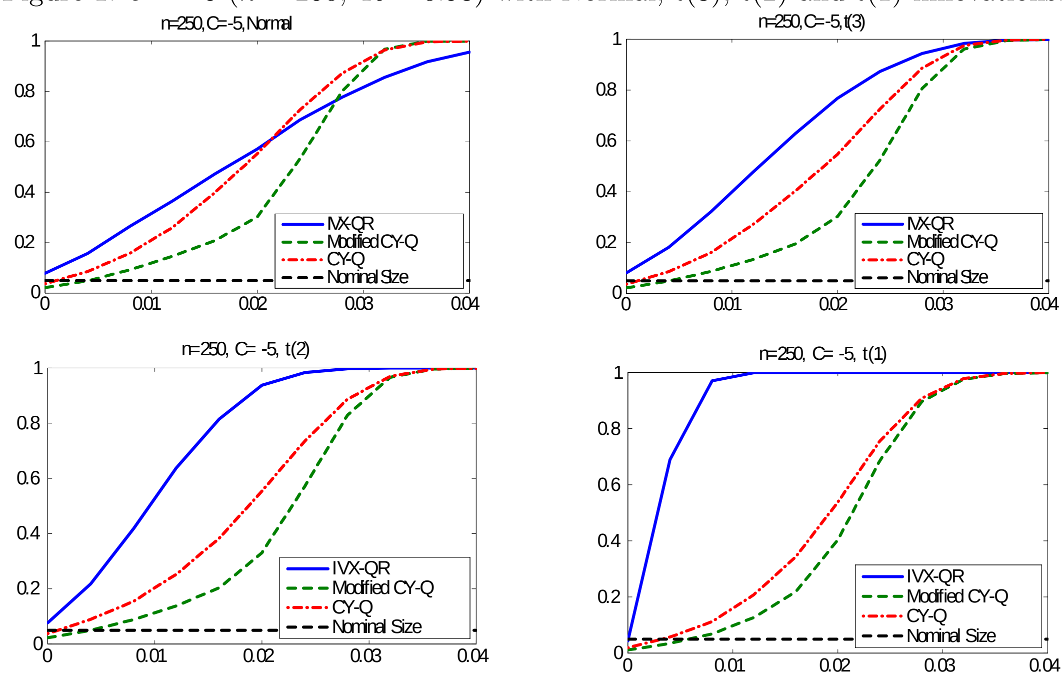


Figure 2: $c = -25$ ($n = 250$, $R = 0.9$) with Normal t(3), t(2) and t(1) innovations.

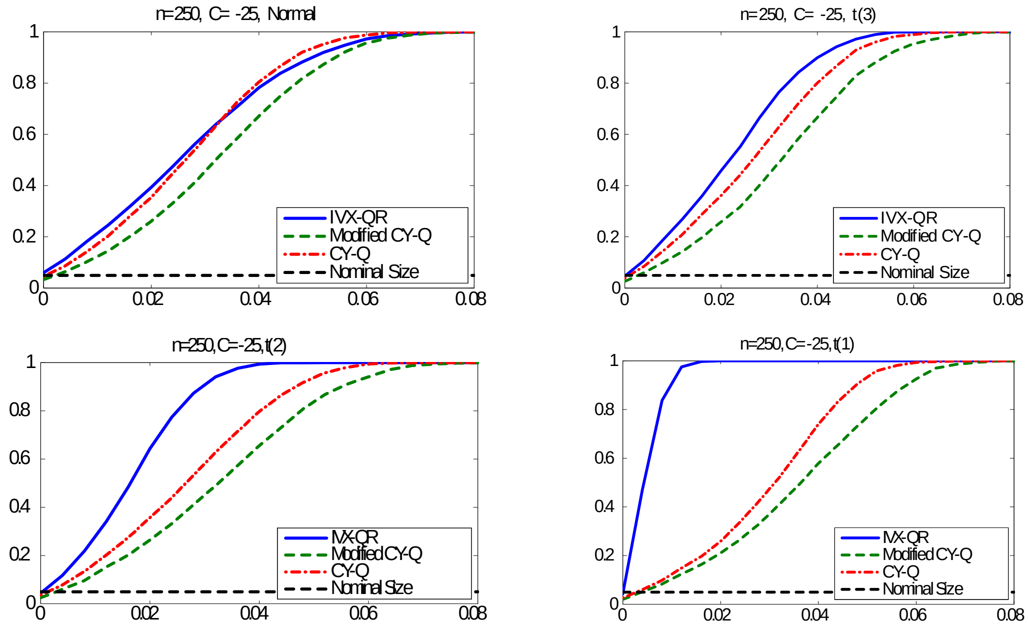
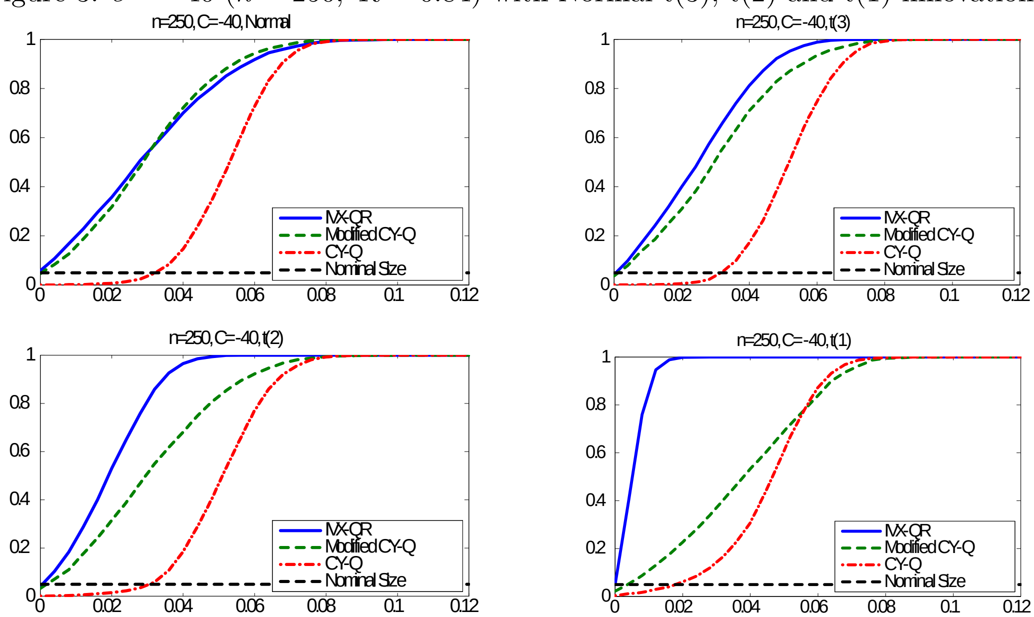


Figure 3: $c = -40$ ($n = 250$, $R = 0.84$) with Normal t(3), t(2) and t(1) innovations.



7 Appendix

The main proofs are given in this Appendix, while the proofs for Lemmas are collected in Lee (2014).

7.1 Proofs for Section 2.2

The following lemma provides the asymptotics of the processes driving the limit theory of $\hat{\beta}_{\tau}^{QR}$.

Lemma 7.1 *1.*

2. $G^x_{\tau,n} := D_n^{-1}\sum_{t=1}^n X_{t-1}\psi_\tau(u_{0t\tau}) \Longrightarrow G^x_\tau$, *where*

$$G^x_\tau = \begin{cases} N\left(0, \tau(1-\tau)\begin{bmatrix} 1 & 0 \\ 0 & \Omega_{xx} \end{bmatrix}\right) & \text{for (I0)}, \\ N\left(0, \tau(1-\tau)\begin{bmatrix} 1 & 0 \\ 0 & V_{xx} \end{bmatrix}\right) & \text{for (MI)}, \\ \begin{bmatrix} B_{\psi_\tau}(1) \\ \int J^c_x(r) dB_{\psi_\tau} \end{bmatrix} & \text{for (I1)}, \\ MN\left[0, \tau(1-\tau)\begin{bmatrix} 1 & 0 \\ 0 & \tilde{V}_{xx} \end{bmatrix}\right], & \text{for (ME)}, \end{cases}$$

with $V_{xx} = \int_0^\infty e^{rC}\Omega_{xx}e^{rC}dr$, $\tilde{V}_{xx} = \int_0^\infty e^{-rC}Y_C Y_C' e^{-rC}dr$ *and* $Y_C \equiv N\left(0, \int_0^\infty e^{-rC}\Omega_{xx}e^{-rC}dr\right)$.

3. $M^x_{\beta_\tau,n} := D_n^{-1}\sum_{t=1}^n f_{u_{0t\tau},t-1}(0)X_{t-1}X_{t-1}'D_n^{-1} \Longrightarrow M^x_{\beta_\tau}$, *where*

$$M^x_{\beta_\tau} = \begin{cases} f_{u_{0\tau}}(0)\begin{bmatrix} 1 & 0 \\ 0 & \Omega_{xx} \end{bmatrix} & \text{for (I0)}, \\ f_{u_{0\tau}}(0)\begin{bmatrix} 1 & 0 \\ 0 & V_{xx} \end{bmatrix} & \text{for (MI)}, \\ f_{u_{0\tau}}(0)\begin{bmatrix} 1 & \int J^c_x(r)' \\ \int J^c_x(r) & \int J^c_x(r)J^c_x(r)' \end{bmatrix} & \text{for (I1)}, \\ f_{u_{0\tau}}(0)\begin{bmatrix} 1 & 0 \\ 0 & \tilde{V}_{xx} \end{bmatrix} & \text{for (ME)}, \end{cases}$$

and convergence in probability holds for (I0) and (MI) cases.

Proof of Theorem 2.1. As in Xiao (2009, Proof of Theorem 1), we can linearize (2.8) in terms of an arbitrary centred quantity $D_n^{-1}(\hat{\beta}_\tau - \beta_\tau)$ using Knight's identity (Knight, 1989). Note that (2.8) is a convex minimization. Using the convexity lemma (Pollard, 1991) we can take the distributional limit of the linearized (2.8) first, and then minimize to get:

$$D_n^{-1}(\hat{\beta}^{QR}_\tau - \beta_\tau) = (M^x_{\beta_\tau,n})^{-1}G^x_{\tau,n} + o_p(1).$$

The results of Theorem 2.1 now follow from Lemma 7.1. ■

7.2 Proofs of IVX-QR Asymptotics: Section 3.2 and 3.3

When x_{t-1} belongs to (I0) or (ME), the limit theory for IVX-QR estimator $\hat{\beta}^{IVXQR}_{1,\tau}$ is identical to that of the ordinary QR estimator $\hat{\beta}_{1,\tau}$ in Theorem 2.1. In general, $\tilde{z}_{t-1}$ reduces the persistence of x_{t-1} when x_{t-1} is more persistent than $\tilde{z}_{t-1}$ ($\delta < \alpha$), except (ME) case. If x_{t-1} is less persistent than $\tilde{z}_{t-1}$ ($\delta > \alpha$) then original persistence of x_{t-1} is maintained, hence the ordinary QR limit theory is achieved. When x_{t-1} is (ME), the remainder term of IVX dominates the asymptotics, see PL for the mean regression framework. I confirm the same results in QR here.

The following lemma provides probability and distributional limit of processes driving the asymptotic behavior of IVX-QR estimators for (I0), (MI) and (I1) cases.

Lemma 7.2 *1.* $G_{\tau,n} := \sum_{t=1}^{n} \tilde{Z}_{t-1,n}\psi_{\tau}(u_{0t\tau}) \Longrightarrow G_{\tau} \equiv \begin{cases} N(0,\tau(1-\tau)\Omega_{xx}) & \textit{for (I0)}, \\ N(0,\tau(1-\tau)V_{cxz}) & \textit{for (MI) and (I1)}, \end{cases}$

2. $M_{\gamma_\tau,n} := \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\,\tilde{Z}_{t-1,n}\tilde{Z}'_{t-1,n} \to^{p} M_{\gamma_\tau} \equiv \begin{cases} f_{u_{0\tau}}(0)\,\Omega_{xx}, & \textit{for (I0)}, \\ f_{u_{0\tau}}(0)\,V_{cxz}. & \textit{for (MI) and (I1)}, \end{cases}$

3. $M_{\beta_\tau,n} := \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\,\tilde{Z}_{t-1,n}X'_{t-1,n} \Longrightarrow M_{\beta_\tau} \equiv \begin{cases} f_{u_{0\tau}}(0)\,\Omega_{xx}, & \textit{for (I0)}, \\ f_{u_{0\tau}}(0)\,\Psi_{cxz}, & \textit{for (MI) and (I1)}, \end{cases}$

To prove Theorem 3.1, I introduce a version of empirical process. Let $\epsilon \in \mathbb{R}^K$ and

$$G_n(\epsilon) = n^{-(1+\delta)/2}\sum_{t=1}^{n} \tilde{z}_{t-1}\left\{\psi_{\tau}\left(u_{0t\tau} - \epsilon' x_{t-1}\right) - E_{t-1}\left[\psi_{\tau}\left(u_{0t\tau} - \epsilon' x_{t-1}\right)\right]\right\}.$$

I focus (MI) and (I1) cases here and assume $0 < \delta < \min(\alpha, 1)$ for documentation purpose; the case of $\alpha \in (0,\delta)$ will be analogous with $n^{-(1+\alpha)/2}$ hence omitted. (I0) case is again standard, so omitted. Proof for (ME) predictors will be discussed below (Lemma 7.4).

The stronger normalizer $n^{-(1+\delta)/2}$ (than $n^{-1/2}$) stabilizes the stronger signal strength of $\tilde{z}_{t-1}$ and x_{t-1}, and the conditional expectation $E_{t-1}[\cdot]$ (rather than unconditional expectation) avoids the nonstationarity problem. Thus, the stochastic equicontinuity proof of Bickel (1975) with iid regressors can be modified accordingly. In fact, $\tilde{z}_{t-1}$ and ψ_τ satisfy condition G and C_1 of Bickel (1975) respectively, hence the analogy of Lemma 4.1 of Bickel (1975) holds.

Lemma 7.3 *For a generic constant* $C > 0$,

$$\sup\left\{\|G_n(\epsilon) - G_n(0)\| : \|\epsilon\| \le n^{(1+\delta)/2}C\right\} = o_p(1)$$

Proof of Theorem 3.1. 1. Since I have confirmed the uniform approximation Lemma 7.3, the standard result for the extremum estimation with non-smooth criterion function (e.g., Pakes and Pollard, 1989) holds with a stronger normalization $n^{-(1+\delta)/2}$. Hence, we can show $(\hat{\beta}_{1,\tau}^{IVXQR} - \beta_{1,\tau}) = O_p(n^{-(1+\delta)/2})$. Let $\hat{\beta}_{1,\tau} = \hat{\beta}_{1,\tau}^{IVXQR}$ within this proof.

Let $\hat{\epsilon}_\tau = (\hat{\beta}_{1,\tau} - \beta_{1,\tau})$, then from (3.7)

$$\begin{aligned}
o_p(1) &= n^{-(1+\delta)/2}\sum_{t=1}^{n}\tilde{z}_{t-1}\left\{\psi_\tau(u_{0t\tau} - (\hat{\beta}_{1,\tau} - \beta_{1,\tau})' x_{t-1})\right\} \\
&= n^{-(1+\delta)/2}\sum_{t=1}^{n}\tilde{z}_{t-1}\left\{\psi_\tau\left(u_{0t\tau} - \hat{\epsilon}_\tau{}' x_{t-1}\right) - E_{t-1}\left(\psi_\tau\left(u_{0t\tau} - \hat{\epsilon}_\tau{}' x_{t-1}\right)\right) - \psi_\tau\left(u_{0t\tau}\right) + E_{t-1}\left(\psi_\tau\left(u_{0t\tau}\right)\right)\right\} \\
&\quad + n^{-(1+\delta)/2}\sum_{t=1}^{n}\tilde{z}_{t-1}E_{t-1}\left(\psi_\tau\left(u_{0t\tau} - \hat{\epsilon}_\tau{}' x_{t-1}\right)\right) + n^{-(1+\delta)/2}\sum_{t=1}^{n}\tilde{z}_{t-1}\left\{\psi_\tau\left(u_{0t\tau}\right)\right\} \\
&= n^{-(1+\delta)/2}\sum_{t=1}^{n}\tilde{z}_{t-1}E_{t-1}\left(\psi_\tau\left(u_{0t\tau} - \hat{\epsilon}_\tau{}' x_{t-1}\right)\right) + n^{-(1+\delta)/2}\sum_{t=1}^{n}\tilde{z}_{t-1}\left\{\psi_\tau\left(u_{0t\tau}\right)\right\} + o_p(1),
\end{aligned}$$

With notation of embedded normalizers,

$$o_p(1) = \sum_{t=1}^{n}\left\{\tilde{Z}_{t-1,n}\psi_\tau\left(u_{0t\tau}\right) + \tilde{Z}_{t-1,n}E_{t-1}\left(\psi_\tau\left(u_{0t\tau} - \hat{\epsilon}_\tau{}' x_{t-1}\right)\right)\right\}, \tag{7.1}$$

and $E_{t-1}\left(\psi_\tau\left(u_{0t\tau} - \hat{\epsilon}_\tau{}' x_{t-1}\right)\right)$ can be expanded around $\epsilon_\tau = 0$ ($\beta_1 = \beta_1(\tau)$), hence

$$E_{t-1}\left[\psi_\tau\left(u_{0t\tau} - \hat{\epsilon}_\tau{}' x_{t-1}\right)\right] = E_{t-1}\left[\psi_\tau\left(u_{0t\tau} - \epsilon_\tau{}' x_{t-1}\right)\right]\Big|_{\epsilon_\tau=0} + \left.\frac{\partial E_{t-1}\left[\psi_\tau\left(u_{0t\tau} - \epsilon_\tau{}' x_{t-1}\right)\right]}{\partial \epsilon_\tau'}\right|_{\epsilon_\tau=0}\hat{\epsilon}_\tau + o_p(\hat{\epsilon}_\tau)$$

where

$$E_{t-1}\left[\psi_\tau\left(u_{0t\tau} - \epsilon_\tau{}' x_{t-1}\right)\right] = \tau - E_{t-1}\left[1\left(u_{0t\tau} < \epsilon_\tau{}' x_{t-1}\right)\right] = \tau - \int_{-\infty}^{\epsilon_\tau{}' x_{t-1}} f_{u_{0t\tau},t-1}(s)ds$$

hence

$$\left.\frac{\partial E_{t-1}\left[\psi_\tau\left(u_{0t\tau} - \epsilon_\tau{}' x_{t-1}\right)\right]}{\partial \epsilon_\tau'}\right|_{\epsilon_\tau=0} = -x_{t-1}' f_{u_{0t\tau},t-1}(0),$$

thus

$$E_{t-1}\left[\psi_\tau\left(u_{0t\tau} - \hat{\epsilon}_\tau{}' x_{t-1}\right)\right] = -x_{t-1}' f_{u_{0t\tau},t-1}(0)\hat{\epsilon}_\tau + o_p(1).$$

Putting it back to (7.1),

$$o_p(1) = G_{\tau,n} + \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\tilde{Z}_{t-1,n}X_{t-1,n}' n^{(1+\delta)/2}\left(\hat{\beta}_{1,\tau} - \beta_{1,\tau}\right),$$

therefore,

$$n^{(1+\delta)/2}\left(\hat{\beta}_{1,\tau} - \beta_{1,\tau}\right) = \left(M_{\beta_\tau,n}\right)^{-1} G_{\tau,n} + o_p(1),$$

and the results of Theorem 3.1 for (MI)-(I1) cases follow from Lemma 7.2. ■

Lemma 7.4 *If x_{t-1} belongs to (ME), then* $n^{-(\alpha\wedge\delta)}R_n^{-n}\sum_{t=1}^{n}\tilde{z}_{t-1}\psi_\tau\left(u_{0t\tau}\right) \Longrightarrow CC_{z\alpha\delta} \times N\left(0, \tau(1-\tau)\tilde{V}_{xx}\right),$

$$and \left\{ \begin{array}{c} \frac{1}{n^{\alpha+(\alpha\wedge\delta)}} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0) R_n^{-n} \tilde{z}_{t-1} x_{t-1}' R_n^{-n} \\ \frac{1}{n^{\alpha+(\alpha\wedge\delta)}} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0) R_n^{-n} \tilde{z}_{t-1} \tilde{z}_{t-1}' R_n^{-n} \end{array} \right\} \Longrightarrow f_{u_{0\tau}}(0) \tilde{V}_{xx} \times CC_{z\alpha\delta}, \ where$$

$$C_{z\alpha\delta} := \left\{ \begin{array}{cc} -C_z^{-1}, & if\ \delta < \alpha \\ C^{-1}, & if\ \alpha < \delta \\ (C - C_z)^{-1}, & if\ \alpha = \delta \end{array} \right\}$$

Proof. The result directly follows from the proof of Lemma 2.4 in Phillips and Lee (2014), by replacing u_{0t} with $\psi_\tau(u_{0t\tau})$. Thus, the result in Theorem 3.1 for (ME) case follows similarly. ∎

Proof of Theorem 3.2. Note that

$$y_{t\tau} - \gamma_1' \tilde{z}_{t-1} = y_{t\tau} - (\gamma_1 - \beta_{1,\tau})' \tilde{z}_{t-1} - \beta_{1,\tau}' \tilde{z}_{t-1}$$
$$= u_{0t\tau} - (\gamma_1 - \beta_{1,\tau})' \tilde{z}_{t-1} + \beta_{1,\tau}'(x_{t-1} - \tilde{z}_{t-1}) = u_{0t\tau}^* - (\gamma_1 - \beta_{1,\tau})' \tilde{z}_{t-1}$$

where $u_{0t\tau}^* = u_{0t\tau} + \beta_{1,\tau}'(x_{t-1} - \tilde{z}_{t-1})$. Following the proof of Theorem 2.1, it is straightforward to show that

$$n^{(1+(\alpha\wedge\delta))/2}(\hat{\gamma}_{1,\tau}^{IVXQR} - \beta_{1,\tau}) = (M_{\gamma_\tau,n})^{-1} G_{\tau,n}^* + o_p(1),$$

where $G_{\tau,n}^* = \sum_{t=1}^n \tilde{Z}_{t-1,n} \psi_\tau(u_{0t\tau}^*)$,and it is clear that $G_{\tau,n}^* = G_{\tau,n}$ under $H_0 : \beta_{1,\tau} = 0$, leading to

$$n^{\frac{1+(\alpha\wedge\delta)}{2}}(\hat{\gamma}_{1,\tau}^{IVXQR} - \beta_{1,\tau}) = (M_{\gamma_\tau,n})^{-1} G_{\tau,n} + o_p(1) \Longrightarrow N\left(0, \tau(1-\tau) f_{u_{0\tau}}(0)^{-2} V_{cxz}^{-1}\right).$$

∎

8 Acknowledgements

This paper is the part of my Ph.D dissertation at Yale University. I am deeply indebted to my advisor Peter Phillips for his support and guidance. I am also grateful to Donald Andrews, Timothy Armstrong, Xiaohong Chen, Timothy Christensen, Yuichi Kitamura, Roger Koenker, Tassos Magdalinos, Nour Meddahi, Taisuke Otsu, Byoung Gun Park, Woong Yong Park, Steve Portnoy, Minkee Song and seminar participants at Yale, Washington, Exeter, Albany, Maryland, NUS, Toulouse and UIUC. I also thank two anonymous referees, associate editor and coeditor for constructive suggestions which improved this paper. All remaining errors are mine.

9 References

Bickel, P.J (1975). "One-Step Huber Estimates in the Linear Model," *Journal of the American Statistical Association*, 70(350), pp.428-434.

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

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

Cavanagh, C., G. Elliott. and J.Stock (1995). "Inference in models with nearly integrated regressors," *Econometric Theory*, 11(05), 1131-1147.

Cenesizoglu, T and A. Timmermann (2008). "Is the distribution of stock returns predictable? Unpublished Manuscript, HEC Montreal and UCSD.

Chernozhukov, V (2005). "Extremal quantile regression," *Annals of Statistics*, 33, 806-839.

Chernozhukov, V. and I. Fernandez-Val (2011). "Inference for extremal conditional quantile models, with an application to market and birthweight risks," *Review of Economic Studies*, 78(2), 559-589.

Davis, R. A., & Mikosch, T. (2009). "The extremogram: a correlogram for extreme events," *Bernoulli*, 15(4), 977-1009.

Elliott, G. and J.H. Stock (1994). "Inference in time series regression when the order of integration of a regressor is unknown," *Econometric Theory*, 10(3-4), 672-700.

Fama, E. and K. French (1993). "Common risk factors in the returns on stocks and bonds, " *Journal of Financial Economics*, 33(1), 3-56.

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

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

Han, H., Linton, O., Oka, T., & Whang, Y. J. (2014). "The cross-quantilogram: measuring quantile dependence and testing directional predictability between time series". arXiv preprint arXiv:1402.1937.

Jansson, M. and M. Moreira (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, 261-278.

Koenker, R (2005). Econometric Society Monographs: *Quantile Regression*. Cambridge Press.

Koenker, R and G. Basset (1978). "Regression quantiles," *Econometrica* 46, 33-49.

Kostakis, A., A. Magdalinos and M. Stamatogiannis (2012). "Robust econometric inference for stock return predictability," Unpublished Manuscript, University of Nottingham.

Lee, J.H. (2014). Online Supplement to "Predictive Quantile Regression with Persistent Covariates: IVX-QR Approach". Available at https://sites.google.com/site/jihyung412/research.

Linton, O., & Whang, Y. J. (2007). "The quantilogram: With an application to evaluating directional predictability," *Journal of Econometrics*, 141(1), 250-282.

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

Maynard A., K. Shimotsu and Y. Wang (2011). "Inference in predictive quantile regressions," Unpublished Manuscript.

Mikusheva, A. (2007). "Uniform inference in autoregressive models," *Econometrica*, 75(5), 1411-1452.

Pakes, A. and D. Pollard (1989). "Simulation and the asymptotics of optimization estimators," *Econometrica* 57(5), pp.1027-1057.

Phillips, P. C. B. (1995). "Fully modified least squares and vector autoregression," *Econometrica* 63(5), 1023–1078.

Phillips, P. C. B (2014), "On confidence Intervals for autoregressive roots and predictive regression," *Econometrica*, 82(3), 1177-1195.

Phillips, P. C. B. (1989), "Partially identified econometric models," *Econometric Theory*, 5(2), pp. 181-240

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

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

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

Phillips, P. C. B and J.H. Lee (2014), "Robust econometric inference with mixed integrated and mildly explosive regressors," Unpublished Manuscript.

Phillips, P. C. B. and T. Magdalinos (2007), "Limit theory for moderate deviations from a unit root," *Journal of Econometrics* 136, 115-130.

Phillips, P. C. B. and V. Solo (1992). "Asymptotics for linear processes", *The Annals of Statistics*, pp. 971-1001.

Pollard, D. (1991)."Asymptotics for least absolute deviation regression estimators," *Econometric Theory*, 7(2), pp. 186-199.

Stock, J. (1991). "Confidence intervals for the largest autoregressive root in US macroeconomic time series," *Journal of Monetary Economics*, 28(3), 435-459.

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