Predictive Quantile Regression with Persistent Covariates*

Ji Hyung Lee†
Department of Economics
Yale University

(Job Market Paper)

November 17, 2012

Abstract

This paper develops econometric methods for inference and prediction in quantile regression (QR) allowing for persistent predictors and possible heavy-tailed innovations. Conventional QR econometric techniques lose their validity when predictors are highly persistent as is the case in much of the empirical finance literature where typical predictors include variables such as dividend-price and earnings-price ratios. I adopt and extend a recent 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 predictors while preserving discriminatory power. The new estimator has a mixed normal limit theory regardless of the degree of persistence in the multivariate predictors. A new computationally attractive testing method for quantile predictability is suggested to simplify implementation for applied work. IVX-QR inherits the robust properties of QR and simulations confirm its substantial superiority over existing methods under thick-tailed errors. The new methods are employed to examine predictability of US stock returns at various quantile levels. The empirical findings confirm greater forecasting capability at quantiles away from the median and suggest potentially improved forecast models for stock returns.

Keywords: Heavy-tailed errors, Instrumentation, IVX filtering, Local to unity, Multivariate predictors, Nonstationary, Predictive regression, Quantile regression, Robustness, Unit root.

JEL classification: C22

---

*I am deeply indebted to my advisor Peter Phillips for his endless support and guidance. I am also grateful to Donald Andrews, Timothy Armstrong, Xiaohong Chen, Hiroaki Kaido, Yuichi Kitamura, Taisuke Otsu and Woong Yong Park for valuable suggestions. Comments made by Yoosoon Chang, David Childers, Timothy Christensen, Byoung Gun Park, Zhentao Shi, Minkee Song and Yale econometrics seminar participants are greatly appreciated. All remaining errors are mine.

†Email: jihyung.lee@yale.edu. Tel: 1 203 859 0689. Comments are welcome.

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. Forward premium regressions and consumption growth regressions are other commonly occurring forecast models that similarly present many empirical puzzles. A central econometric issue in all 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 explored and developed econometric methods for correcting this distortion and validating inference. A recent review of this research is given in Phillips and Lee (2012a, 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 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 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 such as asset returns. 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 I develop for 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 stronger tail dependence. 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 (2009b). 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. A computationally attractive testing method for quantile predictability is developed 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, significant evidence that dividend-price and dividend-payout ratios have predictive ability for lower quantiles of stock returns is provided by our quantile IVX regressions, while T-bill rate and default yields show predictive ability over upper quantiles of subsequent returns. 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 dividend-price ratio and T-bill rate, or the latter with book-to-market ratio are shown to predict almost all stock return quantiles considered over the 1927-2005 period. The forecasting capability of the combination of book-to-market ratio and T-bill rate even remains strong 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 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 regressor in predictive QR by extending the limit theory of Xiao (2009) to a near-integrated regressor case. The last two papers can be classified into predictive QR literature since they focused on the prediction of stock return quantiles from lagged financial variables.

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 most extensive range of predictor persistence ever studied, from stationary predictors through 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. Since the degree of persistence is always imprecisely determined in practical work, uniformity is a significant empirical advantage. Second, the IVX-QR methods validate inference under multiple persistent predictors while 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. 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 finance and macroeconomics.

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 contains simulation results and Section 5 illustrates the empirical examples. Section 6 concludes, while lemmas, proofs and further technical arguments are given in the Appendix.

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, $\mathcal{F}_t$ 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 $C = diag\left(c_1, c_2, ..., c_K\right)$ represents persistence in the multiple predictors of unknown degree. I allow for more general degrees of persistence in the predictors than the existing literature. In particular, x_t can belong to any of the following persistence categories:

(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 $c_i \in (0, \infty)$, $\forall i$.

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:

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

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

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

where $B = (B_0', B_x')'$ is vector Brownian motion (BM). In this instance, 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 introduce a linear predictive QR model. Given the natural filtration $\mathcal{F}_t = \sigma\{u_j = \left(u_{0j}, u_{xj}'\right)', j \le t\}$, 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 given the information $\mathcal{F}_{t-1}$

$$\Pr\left(y_t \le Q_{y_t}(\tau|\mathcal{F}_{t-1})\,|\mathcal{F}_{t-1}\right) = \tau \in (0,1). \tag{2.7}$$

The model (2.6) may be considered a generalization of the predictive regression model (2.1), as it analyzes other quantile predictability as well as the central quantile of y_t. Under the assumption of symmetrically distributed errors, (2.6) with $\tau = 0.5$ reduces to (2.1) since $Q_{y_t}(0.5|\mathcal{F}_{t-1}) = E(y_t|\mathcal{F}_{t-1})$. Letting the QR innovation $u_{0t\tau} := y_t - Q_{y_t}(\tau|\mathcal{F}_{t-1}) = y_t - \beta_{0,\tau} - \beta_{1,\tau}' x_{t-1}$, one obtains

$$\begin{aligned} y_t &= \beta_{0,\tau} + \beta_{1,\tau}' x_{t-1} + u_{0t\tau} \\ &= \beta_\tau' X_{t-1} + u_{0t\tau}, \end{aligned} \tag{2.8}$$

where $X_{t-1} = (1, x_{t-1}')'$, and

$$Q_{u_{0t\tau}}(\tau|\mathcal{F}_{t-1}) = 0 \tag{2.9}$$

from (2.7). In fact, (2.9) is the only moment condition so innovation with infinite mean or variance is allowed[1]. The robustness of QR to thick-tailed errors enables better econometric inference than mean regression for financial applications using heavy-tailed data. To see the relation of (2.1) and (2.8), recenter the innovation of (2.1) as $u_{0t\tau} = u_{0t} - F_{u_0}^{-1}(\tau)$, then (2.9) follows, and

$$y_t = \beta_{0,\tau} + \beta_1' x_{t-1} + u_{0t\tau} \text{ and } \beta_{0,\tau} = \beta_0 + F_{u_0}^{-1}(\tau),$$

which is the construction of Xiao (2009).

By defining a piecewise derivative of the loss function in the QR $\psi_\tau(u) = \tau - 1(u < 0)$ (see below), it is easy to show the QR "induced" innovation $\psi_\tau(u_{0t\tau}) \sim mds(0, \tau(1-\tau))$, and the

[1] The functional CLT in (2.4) holds under the finite second moment condition as in (2.3), and we will need α-stable limit laws for infinite mean/variance innovation cases. To stay focused, I only consider limit theory with finite variance innovations here but IVX-QR limit theory can be extended to the case of infinite mean/variance innovations, as we will show in later work. The self-normalized IVX-QR estimator shows extremely good normal approximations under infinite mean/variance errors in the simulation section.

following functional law holds

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor}\left[\begin{array}{c}\psi_{\tau}\left(u_{0t\tau}\right)\\ u_{xt}\end{array}\right]\Longrightarrow\left[\begin{array}{c}B_{\psi_{\tau}}(r)\\ B_{x}(r)\end{array}\right]=BM\left[\begin{array}{cc}\tau\left(1-\tau\right) & \Sigma_{\psi_{\tau}x}\\ \Sigma_{x\psi_{\tau}} & \Omega_{xx}\end{array}\right]. \tag{2.10}$$

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* $f_{u_{0\tau}}(0)=E\left[f_{u_{0t\tau},t-1}(0)\right]>0$,

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

(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.,*

$$\left|f_{u_{0t\tau},t-1}(\epsilon)\right|<\infty \ \textit{w.p.1}$$

where ϵ *can be arbitrarily close to zero.*

Remark 2.1 *Assumption 2.1-(i) is not restrictive considering that an mds structure is commonly imposed on* u_{0t} *(hence* $u_{0t\tau}$*) in the predictive regression literature. Assumption 2.1-(ii) is a standard technical condition used in the QR literature, enabling 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 Nonstationary Quantile Regression

This subsection 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:

$$\hat{\beta}_{\tau}^{QR}=\arg\min_{\beta}\sum_{t=1}^{n}\rho_{\tau}\left(y_t-\beta' X_{t-1}\right) \tag{2.11}$$

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:

$$D_n:=\begin{cases}\sqrt{n}I_{K+1} & \text{for (I0)},\\ diag(\sqrt{n},n^{\frac{1+\alpha}{2}}I_K) & \text{for (MI)},\\ diag(\sqrt{n},nI_K) & \text{for (I1)},\\ diag(\sqrt{n},n^{\alpha}R_n^n) & \text{for (ME)}.\end{cases} \tag{2.12}$$

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

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

where $V_{xx}=\int_0^\infty e^{rC}\Omega_{xx}e^{rC}dr$, $\tilde{V}_{xx}=\int_0^\infty e^{-rC}Y_CY_C'e^{-rC}dr$ *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

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

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) $dB_{\psi_\tau}=dB_{\psi_\tau.x}+\left(\frac{\Sigma_{\psi_\tau x}}{\Sigma_{xx}}\right)dB_x$ we have

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

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 $\Sigma_{\psi_\tau.x} = \tau(1-\tau) - \Sigma_{xx}^{-1}\Sigma_{\psi_\tau x}^2$ and $\Sigma_{\psi_\tau x} = E\left[1\left(u_{0t\tau} < 0\right)u_{xt}\right] = E\left[1\left(u_{0t} < F_{u_0}^{-1}(\tau)\right)u_{xt}\right]$. Now assume a researcher uses the ordinary QR standard error:

$$s.e(\widehat{\beta}_{1,\tau}^{QR}) = \left[\frac{\tau(1-\tau)}{\hat{f}_{u_{0\tau}}(0)^2}\left(\frac{1}{n^2}\sum_{t=1}^{n}\left(x_{t-1}^{\mu}\right)^2\right)^{-1}\right]^{1/2},$$

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:

$$\begin{aligned} t_{\widehat{\beta}_{1,\tau}} &= \frac{\left(\widehat{\beta}_{1,\tau}^{QR} - \beta_{1,\tau}\right)}{s.e(\widehat{\beta}_{1,\tau}^{QR})} \\ &\sim \left[1 - \frac{\Sigma_{\psi_\tau x}^2}{\Sigma_{xx}\tau(1-\tau)}\right]^{1/2} N(0,1) \quad + \left[\frac{\Sigma_{\psi_\tau x}^2}{\Sigma_{xx}\tau(1-\tau)}\right]^{1/2}\frac{\int \bar{I}_x^c dW_x}{\left[\int\left(\bar{I}_x^c\right)^2\right]^{1/2}} \\ &\sim \underbrace{\left[1-\lambda(\tau)^2\right]^{1/2} Z}_{\text{standard inference}} \quad + \underbrace{\lambda(\tau)\,\eta_{LUR}(c)}_{\text{nonstandard distortion}} \end{aligned} \tag{2.13}$$

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

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

Remark 2.2 *As the analytical expression (2.13) shows, the nonstandard distortion becomes greater with (i) smaller* $|c|$ *and (ii) larger* $|\lambda(\tau)|$*. Condition (i) is well known from 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 quantile regression, see Xiao (2009) for strict unit root regressors. Note that:*

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

so the explicit source of distortion from persistence and nonlinear dependence is provided by this analysis. The commonly used lower tail dependence measure is

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

Thus, $\lambda(\tau)$ *is a measure of the dependence between the linear dependence (*$E\left[u_{0t}u_{xt}\right]$*) and tail dependence (*$E\left[1\left(u_{0t} < F_{u_0}^{-1}(\tau)\right)1\left(u_{xt} < F_{u_x}^{-1}(\tau)\right)\right]$*). If we consider a lower quantile prediction, such as* $\tau = 0.05$ *or* 0.1*, the distortion is more affected by tail dependence than by correlation. This is consistent with the recent findings that a precise VaR estimation depends on the tail dependence structure (e.g., Embrechts et al.,1997). The distortion stemming from the interaction between persistence and tail dependence has not been properly addressed in the literature.*

To correct the nonstandard distortion in (2.13), we may consider two solutions. 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 is the main idea of the Bonferroni methods, which are frequently used in predictive mean regression (e.g., Cavanagh et al.,1995; Campbell and Yogo, 2006). In predictive QR, Maynard et al. (2011) employed the same idea. For a single local to unity (I1) predictor, these methods successfully correct the distortion, but lose their validity when predictor persistence belongs to (MI) or (I0) spaces (Phillips, 2012)[2]. In addition, Bonferroni methods are not applicable in the presence of multivariate nonstationary predictors (multiple c_i's). The second solution to correct for nonstandard distortion, which this paper follows, is to use IVX filtering technique (Magdalinos and Phillips, 2009). Methods based on the IVX filtering technique are discussed in the next section. The main advantages of these methods over (currently available) Bonferroni methods are its uniform validity over a much wider range of predictor persistence including (I0) through (ME) and applicability to multiple persistent predictors, as well as a single persistent predictor.

3 IVX-QR Methods

It is convenient to transform the model (2.8) to remove the explicit intercept term:

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

where $y_{t\tau} := y_t - \hat{\beta}_{0,\tau}^{QR}(\tau) = y_t - \beta_{0,\tau} + O_p\left(n^{-1/2}\right)$ 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. Using the pre-estimated $\sqrt{n}$-consistent estimation $\hat{\beta}_{0,\tau}^{QR}(\tau)$ is asymptotically innocuous (see Appendix 7.2). After estimating $\hat{\beta}_{1,\tau}^{IVXQR}$, we may compute $\hat{\beta}_{0,\tau}^{IVXQR}$ for inference on $\beta_{0,\tau}$ as well (see Appendix 7.3).

3.1 IVX Filtering

This subsection 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

[2]The Bonferroni methods based on a uniformly valid CI for c, rather than Stock's CI, may provide validity over (MI) or (I0) spaces. These extensions, however, have not been developed in the predictive regression literature yet.

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 a statistically challenging task. In terms of convergence rate, the first difference reduces the n-rate (for the (I1) case) to the $\sqrt{n}$-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 (2.13).

When $F = I_K$ then $\tilde{z}_t = x_t$ so we may 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 a nonnegligible 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 still suffers from the serious size distortion in (2.13).

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. 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 far 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 $n^{\frac{1+\delta}{2}}$. At the cost of the slight reduction in convergence rate compared to the level data, the filtering achieves the desired size correction. Simulation in Section 4 shows that this cost is not substantial so we have comparable, and sometimes better local power to n-consistent testing such as Q-test of Campbell and Yogo (2006). To summarize

Table 1: 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	$\sqrt{n}$	$n^{\frac{1+\delta}{2}}$

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 the 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, 2012b). 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

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

and $\alpha \wedge \delta = \min(\alpha, \delta)$. I also define the different asymptotic moment matrices for the (MI) and (I1) cases in a unified way:

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

and

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

Considering the conditional moment restriction $Q_{u_{0t\tau}}(\tau|\mathcal{F}_{t-1}) = 0$, or

$$E\left[\tau - 1\left(y_{t\tau} \leq \beta_{1,\tau}{}' x_{t-1}\right) | \mathcal{F}_{t-1}\right] = 0,$$

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*

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

where W_n is an arbitrary weighting matrix and

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

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)

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

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,*

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

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

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

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. When the number of parameter K is large, the procedure (3.6) may be

[3]The kernel density estimation with standard normal kernel functions and a bandwidth choice of Silvermann's rule is used in the simulation and empirical results below.

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 $H_0 : \beta_{1,\tau} = 0$ (significant testing), 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:

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 $H_0 : \beta_{1,\tau} = \frac{b_{1,\tau}}{n^\nu}$ 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 $H_0 : \beta_{1,\tau} = 0$,*

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

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 we could use the following test statistics

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

where $H = \begin{bmatrix} 1 & 0 & \mathbf{0}_{1\times(K-2)} \\ 0 & 1 & \mathbf{0}_{1\times(K-2)} \end{bmatrix}$.

4 Simulation

In this section, I conduct simulations to observe the size and power performance of IVX-QR inference methods. Theorems 3.1 and 3.2 show that a larger IVX persistence (δ) leads to a more efficient test, while a smaller δ generally achieves better size corrections. Using simulations with settings similar to those used in the existing predictive regression literature, I provide guidance for the choice of δ in applied work.

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 $\mu_x = 0$, $R = I_K + \frac{C}{n}$ and

$$u_t = \begin{bmatrix} u_{0t} \\ u_{xt} \end{bmatrix} \sim iid\ F_u\left(0_{(K+1)\times 1}, \Sigma_{(K+1)\times(K+1)}\right). \tag{4.2}$$

The IVX is constructed as (3.2) and the inference uses Proposition 3.2. I normalize $C_z = -5I_K$ and vary δ to show performance according to varying IVX persistence.

4.1 Size and Power Performance with a Single Persistent Predictor

This subsection shows simulations with a single local to unity ($K = 1$, $\alpha = 1$) regressor. Although IVX-QR methods allow testing for predictability at various quantile levels, I focus on the central (median) quantile to compare its performance to that of existing mean predictability tests. In particular, the IVX median test is compared to the methods of Campbell and Yogo (2006; CY-Q) and Kostakis et al (2012; IVX-mean).

For the distribution F_u in (4.2), I first use a normal distribution with the variance

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

with $\phi = -0.95$. This value reflects the realistic error correlation in predictive regressions, such as dividend-price ratio. It also has been frequently used in the literature, such as in Campbell and Yogo (2006) and Jansson and Moreira (2006).

Tables 2 and 3 summarize the size performances of the IVX-QR and CY-Q tests. The empirical size is calculated from the rejection frequency of a one-sided test of $H_0 : \beta_{1,\tau} = 0$ versus $H_1 : \beta_{1,\tau} > 0$ with $\tau = 0.5$. For simplicity, I set $\beta_{0,\tau} = 0$. Note that the CY-Q test uses the null of $H'_0 : E_{t-1}[y_t] = \beta_1 = 0$ instead of H_0 defined above. The assumptions of conditional mean/median zero y_t under H_0 and H'_0 are then exploited to compare these two different types of tests. The introduction of estimated intercepts to relax this assumption is discussed in Appendix 7.5. The nominal size is 0.05 with 2,500 repetitions (S).

Table 2: Finite sample sizes of IVX-QR and CY-Q tests ($n = 100$)

	$R = 1$	$R = 0.98$	$R = 0.9$	$R = 0.8$	$R = 0.5$	$R = 0$
$S = 2500$	$c = 0$	$c = -2$	$c = -10$	$c = -20$	$c = -50$	$c = -100$
$\delta =$	IVXQR	IVXQR	IVXQR	IVX-QR	IVXQR	IVXQR
0.95	0.0724	0.0624	0.0564	0.0456	0.0492	0.0376
0.85	0.0552	0.0584	0.0468	0.0472	0.0412	0.0392
0.75	0.062	0.0572	0.0396	0.0456	0.048	0.0304
0.65	0.0608	0.0556	0.0428	0.0504	0.0324	0.0412
0.55	0.0444	0.0424	0.0456	0.0484	0.0544	0.0344
CY-Q	0.05	0.046	0.0608	0.0792	N/A	N/A

It is evident from Table 2 that the size of IVX-QR is uniformly well controlled in the case of a pure stationary predictor ($R = 0$) to a unit root predictor ($R = 1$), supporting one of the main contributions of this paper. The CY-Q test also shows good size properties for (near) unit root predictors ($R \in [0.9, 1]$), but the performance decays for less persistent predictors. This is because Stock's (1991) CI for R is invalid for (near) stationary regressors, thereby affecting the Bonferroni CI for β_1[4]. I also confirm the robustness of IVX-QR for various values of $\delta \in [0.55, 0.95]$. A similar result holds with $n = 200$.

Table 3: Finite sample sizes of IVX-QR and CY-Q tests ($n = 200$)

	$R = 1$	$R = 0.99$	$R = 0.95$	$R = 0.9$	$R = 0.5$	$R = 0$
$S = 2500$	$c = 0$	$c = -2$	$c = -10$	$c = -20$	$c = -100$	$c = -200$
$\delta =$	IVXQR	IVXQR	IVXQR	IVXQR	IVXQR	IVXQR
0.95	0.0664	0.0616	0.052	0.0492	0.0556	0.042
0.85	0.0716	0.0604	0.0548	0.0564	0.0444	0.0376
0.75	0.0552	0.0652	0.044	0.0504	0.0416	0.0388
0.65	0.0492	0.0492	0.0492	0.0432	0.0544	0.0416
0.55	0.0472	0.0472	0.0448	0.0484	0.0412	0.0388
CY-Q	0.0492	0.0508	0.05	0.05	N/A	N/A

To investigate 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, IVX-mean, and CY-Q tests. For the distribution F_u, I employ normal and t-distributions with four to one (Cauchy) degrees of freedom. As expected, two mean predictability tests dominate IVX-QR under normally distributed errors. For the finite sample performance of the IVX-mean and CY-Q tests under normally distributed innovations, see Kostakis et al. (2012). I report the power performances with t(4)-t(1) innovations to highlight possible improvements when using IVX-

[4] For a detailed explanation of this problem, the reader is referred to recent work by Phillips (2012).

QR in cases of thick-tailed errors. It is widely known that financial asset returns have heavy-tailed distributions (e.g., Cont, 2001). The $c_z = -5$, $\delta = 0.95$ specification is employed because it effectively controls the size. With sample size $n = 200$, I observe the cases of $c \in \{0, -2, -20, -40\}$ reflecting persistent predictors in practice.

Figures 1-4 illustrate the results. For unit root ($\rho = 1$) and near unit root ($\rho = 0.99$) predictors, the test performance rankings with t(3)-t(4) errors are mixed. IVX-QR performs best with mildly persistent predictors, as shown by $\rho = 0.9$ and 0.8 in the simulation. For infinite variance (t(2)) and mean (t(1)) errors, IVX-QR shows the best performance across all scenarios.

Considering that much applied work also uses the intercept term in the regression, IVX-QR with dequantiling, as in (3.1), is compared with the IVX-mean and CY-Q tests. Appendix 7.5 shows the results with the same ranking patterns as above.

In summary, IVX-QR testing with a single persistent predictor is competitive when we have heavy-tailed errors and when the persistence of predictors is close to near stationary (mildly integrated). Because we can obtain information on the tail properties of any given data and partial information, albeit imperfect, on the persistence of predictors, one may decide which test is better to use. All three tests perform well in terms of size and power except for the CY-Q test in cases of mildly integrated predictors. Therefore, several valid methods can test the predictive ability of a single persistent predictor on given financial asset returns. Note that, unlike the CY-Q test, the IVX-mean and IVX-QR tests can 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 subsection.

Figure 1: Power curves of IVX-QR, IVX-mean and CY-Q tests
$c = 0$ ($n = 200$, $R = 1$) with t(4), t(3), t(2) and t(1) innovations.

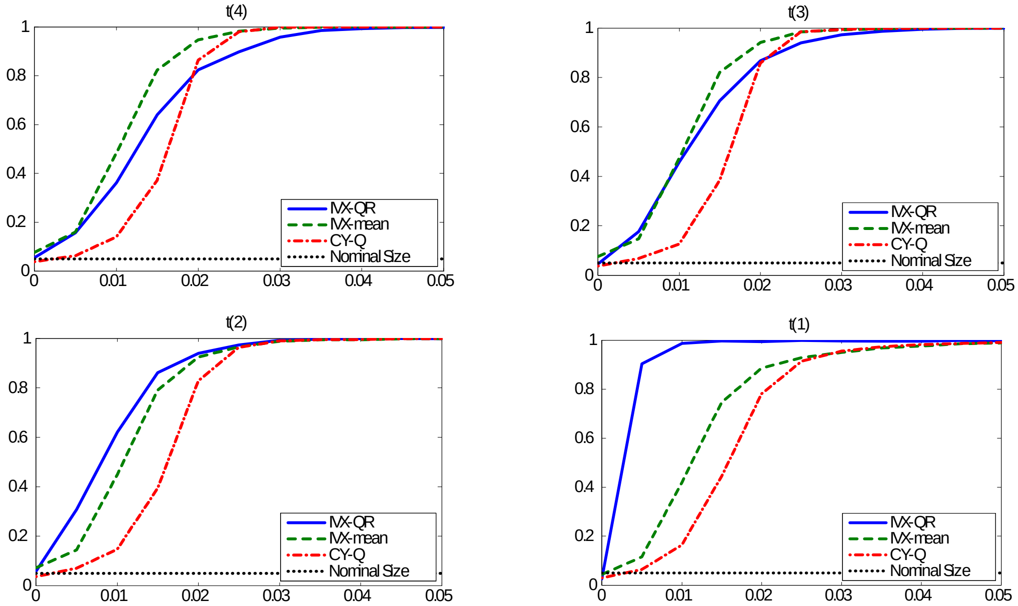


Figure 2: Power curves of IVX-QR, IVX-mean and CY-Q tests
$c = -2$ ($n = 200$, $R = 0.99$) with t(4), t(3), t(2) and t(1) innovations.

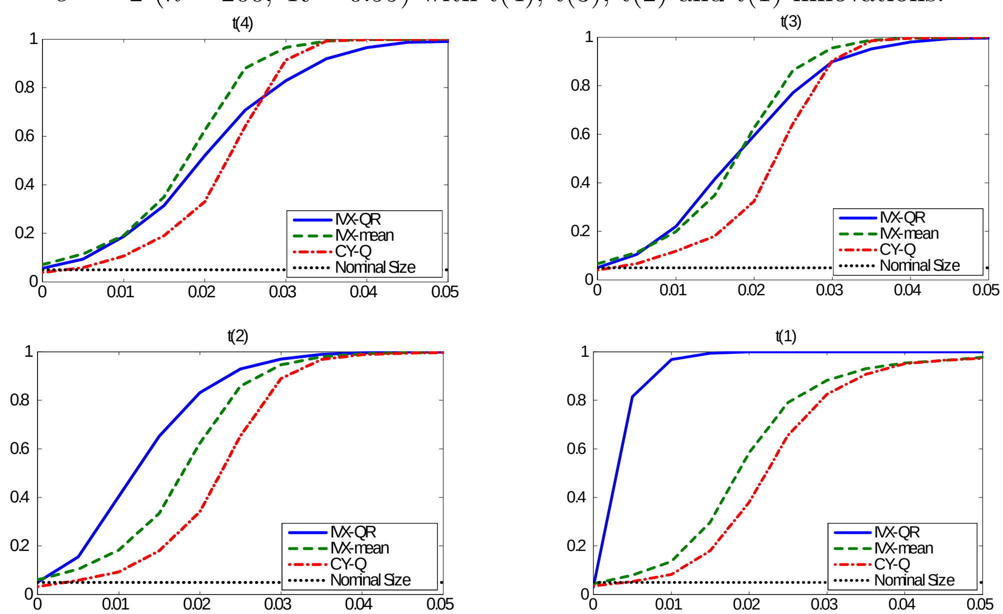


Figure 3: Power curves of IVX-QR, IVX-mean and CY-Q tests
$c = -20$ ($n = 200$, $R = 0.9$) with t(4), t(3), t(2) and t(1) innovations.

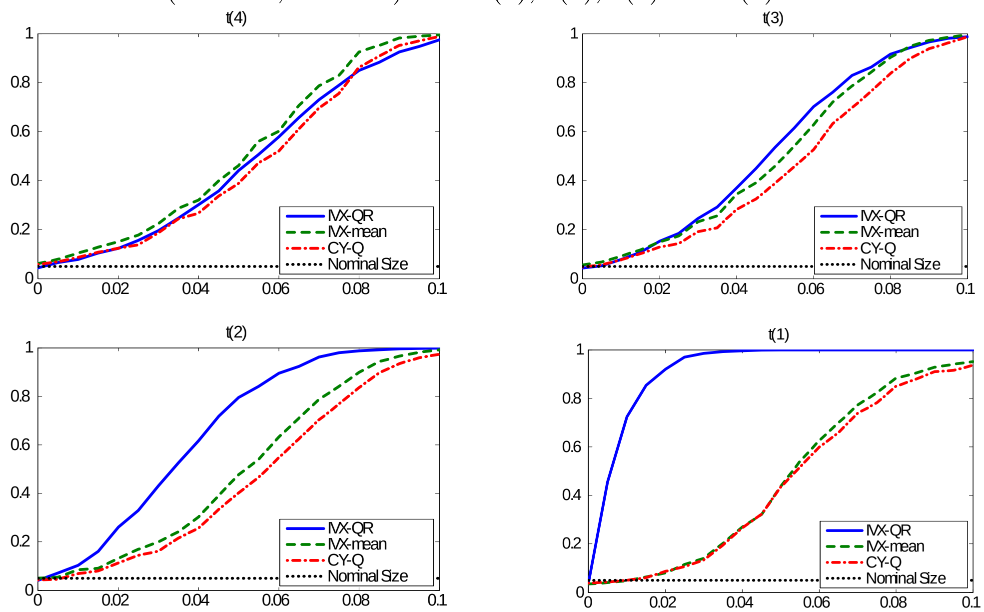
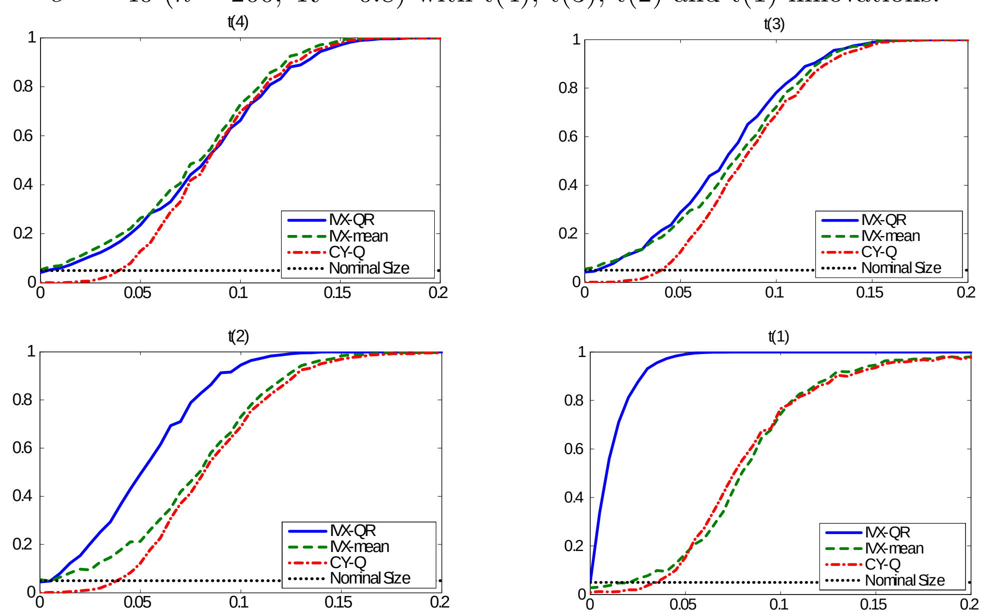


Figure 4: Power curves of IVX-QR, IVX-mean and CY-Q tests
$c = -40$ ($n = 200$, $R = 0.8$) with t(4), t(3), t(2) and t(1) innovations.



4.2 Size Properties of Prediction Tests on Various Quantiles

Despite the vast literature on predictive mean regression, 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 the first valid method to test various quantile predictability of asset returns in the presence of multiple persistent predictors. In this subsection, 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 extreme quantiles with finite sample is a common problem in QR. Large sample sizes are not uncommon in financial applications, and the empirical work in the next section corresponds to one of those applications.

The simulation environment used to test the size properties of various quantile predictions is similar to the earlier subsection, but I now include the estimated intercept to mimic the common practical work. Dequantiling in (3.1) is therefore used for all IVX-QR simulations in this subsection. 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 0.05. I use a cut-off rule of 8% for the reasonable size results because 7-8% is the permissible empirical size level corresponding to a nominal 5% size (see Campbell and Yogo, 2006). The results of these simulations are presented in Tables 4-8 below; the size performances exceeding 8% are shown in bold.

I first investigate the size properties of ordinary QR methods. Table 4 below summarizes the size properties of ordinary QR t-statistics in (2.13) 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 dependence structure of F_u significantly affects the magnitude of the size distortion. For a t-distribution with stronger tail dependence (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 (tail independent). The overall results indicate the invalidity of the ordinary QR technique in the presence of persistent predictors.

The size performances of the IVX-QR methods with $C_z = -5$ and $\delta = 0.5$ are reported in Table 5. Under normally distributed errors, the size corrections are remarkable, confirming the uniform validity of IVX-QR methods at various quantiles. For heavy-tailed errors with stronger tail dependence, the results still hold in all but a few tail cases. Table 6 shows the results with $\delta = 0.6$. In this case, we lose a few more unit root ($c = 0$) cases under stronger tail dependence. Extensive simulation results indicate that the IVX-QR correction methods control sizes remarkably well across all quantiles when δ is less than 0.7. When δ exceeds 0.7, the test tends to over-reject at tails while maintaining good performances at inner quantiles[5].

[5] Simulation results with various choices of δ are readily available upon request.

Table 4: 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**	**9.9**	**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(4) errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c = 0$	**17.2**	**16.6**	**16.7**	**13.9**	**13.1**	**13.2**	**14.6**	**14.1**	**16.7**	**18.0**	**18.3**
$c = -2$	**13.0**	**13.4**	**11.0**	**9.9**	**9.8**	**9.7**	**8.1**	**10.5**	**13.8**	**12.8**	**16.3**
$c = -5$	**12.0**	**10.0**	**10.3**	6.0	7.5	6.4	6.5	**8.5**	**9.8**	**11.6**	**11.2**
$c = -7$	**10.5**	**10.4**	**8.3**	6.8	5.7	6.7	5.6	**8.1**	**8.1**	**10.8**	**10.1**
$c = -70$	**10.0**	6.7	6.9	4.8	4.1	4.6	4.7	4.3	6.1	6.0	**10.1**

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 5: Size Performances(%) of IVX-QR with $\delta = 0.5$ ($n = 700$, $S = 1000$)

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
$c = 0$	7.0	6.0	6.9	4.6	5.1	4.5	6.7	6.0	6.5	5.9	8.0
$c = -2$	7.4	5.6	4.6	4.3	4.8	3.7	5.3	4.0	4.8	5.8	7.5
$c = -5$	6.6	5.5	5.1	4.2	4.8	3.2	3.7	4.5	4.7	5.2	6.8
$c = -7$	5.5	5.6	4.5	4.9	3.6	3.4	3.7	4.1	5.1	4.4	7.3
$c = -70$	4.9	4.8	4.6	4.0	3.1	3.4	3.7	4.1	3.9	4.8	7.1
$c = -700$	6.5	6.8	4.3	6.2	4.3	4.6	5.4	4.8	4.0	4.2	6.5

t(4) errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c = 0$	**9.1**	7.5	6.0	5.9	6.4	4.5	6.0	6.1	6.5	**8.8**	**9.5**
$c = -2$	7.4	5.6	4.0	3.7	3.4	3.8	5.5	5.0	4.9	6.0	**9.8**
$c = -5$	7.5	7.2	5.2	5.0	3.5	3.3	3.4	3.4	5.2	6.0	7.5
$c = -7$	6.5	7.6	5.6	4.9	3.3	3.8	4.8	4.6	5.8	4.8	6.8
$c = -70$	6.5	5.6	5.2	4.1	3.5	3.9	3.2	3.8	5.4	5.6	7.1
$c = -700$	8.0	7.6	4.8	5.4	4.9	3.9	3.9	3.8	6.6	5.7	7.8

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
$c = 0$	**9.8**	**10.2**	7.8	6.4	5.2	5.1	4.8	4.8	6.7	7.6	**11.1**
$c = -2$	7.7	8.3	6.2	4.5	4.4	3.1	4.4	5.5	4.8	7.2	**8.6**
$c = -5$	6.5	7.0	6.0	4.6	3.7	4.1	3.5	3.0	5.2	6.0	6.5
$c = -7$	6.0	5.7	4.9	3.4	3.3	3.7	3.6	4.9	5.2	5.5	7.3
$c = -70$	7.2	7.3	6.1	4.1	4.0	2.8	4.4	5.5	5.0	6.1	5.2
$c = -700$	6.7	6.9	5.8	4.8	4.0	4.4	4.9	4.5	6.0	5.0	6.5

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
$c = 0$	**10.5**	8.9	6.3	5.8	4.7	4.6	5.3	6.4	8.0	5.9	7.7
$c = -2$	7.9	5.9	6.1	3.3	4.5	3.9	3.2	4.1	7.2	6.5	5.6
$c = -5$	5.3	5.9	6.0	6.0	3.7	3.3	5.3	4.2	7.1	6.5	6.2
$c = -7$	6.8	6.4	6.4	3.8	3.3	4.1	4.3	3.7	6.0	**8.4**	7.3
$c = -70$	4.9	7.9	6.6	4.6	3.9	4.3	4.1	4.8	6.3	5.9	5.2
$c = -700$	4.4	5.5	6.2	6.4	4.7	3.8	5.6	5.3	5.2	6.8	3.8

Table 6: Size Performances(%) of IVX-QR with $\delta = 0.6$ ($n = 700$, $S = 1000$)

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
$c = 0$	7.0	7.8	**8.9**	6.7	6.7	6.8	**8.5**	7.9	**8.2**	7.2	**8.9**
$c = -2$	7.6	6.0	5.1	5.0	4.6	4.1	5.1	4.1	5.3	5.7	7.2
$c = -5$	6.6	6.1	4.7	4.3	4.6	3.0	4.7	4.5	5.0	5.4	6.4
$c = -7$	7.1	5.9	4.9	5.3	4.1	4.0	4.1	4.4	4.7	4.9	6.5
$c = -70$	5.1	5.5	4.3	3.3	3.3	3.1	3.5	4.8	4.4	5.0	7.7
$c = -700$	7.4	6.8	4.7	5.7	4.4	4.6	5.5	4.7	3.7	4.1	7.0

t(4) errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c = 0$	**10.8**	**8.4**	**8.2**	7.7	7.6	5.6	6.8	8.0	**9.7**	**9.1**	**9.7**
$c = -2$	8.0	6.7	5.8	4.4	4.4	4.8	5.4	6.2	6.3	6.5	**10.3**
$c = -5$	**8.4**	7.9	5.4	4.5	3.9	3.3	4.4	3.7	4.7	6.2	7.7
$c = -7$	7.5	**8.1**	5.5	5.0	2.5	4.0	4.8	4.8	6.2	5.5	7.4
$c = -70$	6.9	6.4	5.0	4.2	3.7	4.5	3.2	4.1	5.6	6.0	6.9
$c = -700$	8.0	7.2	4.9	5.6	4.7	4.1	3.7	4.1	6.7	6.0	7.7

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
$c = 0$	**10.6**	**11.4**	**9.4**	7.1	7.1	7.2	5.6	7.7	7.7	**9.1**	**10.7**
$c = -2$	**9.1**	7.6	6.0	5.8	4.8	3.9	4.9	5.4	5.7	7.8	**8.2**
$c = -5$	5.8	7.4	5.6	5.1	4.1	3.8	3.5	3.7	5.6	6.9	7.2
$c = -7$	**8.2**	6.5	5.1	4.2	4.3	4.2	3.5	4.3	5.7	5.0	7.8
$c = -70$	7.6	7.4	5.8	3.2	4.0	3.1	4.6	4.6	5.2	5.8	6.7
$c = -700$	7.0	6.8	5.5	3.9	3.9	4.3	5.3	4.4	6.5	5.3	6.5

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
$c = 0$	**13.5**	**9.7**	7.3	6.6	5.6	5.2	6.5	7.6	**9.7**	7.4	**8.8**
$c = -2$	**8.1**	6.6	6.6	4.4	4.3	3.9	3.1	4.7	**8.6**	7.9	5.5
$c = -5$	6.5	7.3	6.4	5.3	4.9	3.8	5.4	4.7	7.1	7.6	7.1
$c = -7$	8.0	7.7	6.2	4.6	3.3	3.8	4.4	4.3	6.7	**8.2**	7.5
$c = -70$	5.1	7.2	6.1	4.5	4.1	3.5	3.9	4.8	6.8	6.0	6.5
$c = -700$	3.8	5.1	5.6	6.2	4.8	3.9	6.5	5.4	5.7	6.8	4.1

I now consider the predictive QR scenario with multiple persistent predictors ($K = 2$). This scenario has rarely been explored but is highly relevant in empirical practice (e.g., dividend-price ratio and Treasury bill rate). The bivariate predictor persistence (c_1, c_2) is again selected from $\{0, -2, -5, -7, -70\}$ and all other simulation settings are the same as before. For the innovation structure, I borrow a calibration technique to avoid lengthy documentation. In the next section, specification with two predictors of book-to-market ratio and Treasury bill rate is shown to predict stock returns at various quantile levels from January 1952 to December 2005. To support the empirical finding, the estimated correlation of the predictive QR application is used:

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

Table 7 shows the size properties of ordinary QR test statistics. The size distortion is larger when there is more than one persistent predictor, which corroborates the benefits of the IVX-QR method's validating inference under multiple manifestations of predictor persistence. Table 8 shows acceptable size results of the IVX-QR tests at various quantile levels. The size correction under normally distributed errors works very well for most quantiles. Performances at tails are good except for a few cases, and results at inner quantiles from 0.1 to 0.9 are satisfactory. Under strong tail dependent innovations (t(4) or t(3) errors), the IVX-QR corrections for multiple persistent predictors at $\tau = 0.05$ or 0.1 are partly acceptable. The IVX-QR correction result for these extreme cases suggests a need for new methods to handle extremal quantiles. One promising potential solution could be the use of a recent development in extremal QR limit theory (e.g., Chernozhukov, 2005; Chernozhukov and Fernandez-Val, 2012), but I leave this for future research.

In summary, IVX-QR methods demonstrate reliable size performances for all relevant specifications with single and multiple persistent predictors, except for a few extreme cases. The practical benefits of IVX-QR inference will be illustrated through empirical examples in the next section.

Table 7: 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
$c_1 = c_2 = 0$	**14.0**	**12.5**	**13.2**	**14.6**	**13.7**	**13.7**	**13.2**	**12.1**	**13.4**	**11.4**	**15.4**
$c_1 = -2$, $c_2 = 0$	**12.7**	**11.5**	**10.1**	**10.9**	**9.2**	**9.8**	**11.6**	**13.0**	**11.4**	**10.5**	**12.8**
$c_1 = -5$, $c_2 = -0$	**13.5**	**10.4**	**9.0**	**8.1**	**9.6**	**6.9**	**8.1**	**8.5**	8.0	**10.1**	**13.2**
$c = -7$, $c_2 = -0$	**9.6**	**9.9**	7.2	**8.4**	7.7	7.4	7.0	7.5	**8.4**	**9.2**	**10.1**
$c_1 = -2$, $c_2 = -2$	**11.3**	**10.4**	**9.3**	**9.2**	**9.5**	**10.5**	**9.6**	**8.8**	**10.1**	**10.4**	**12.6**
$c_1 = -5$, $c_2 = -2$	**10.5**	**9.5**	**8.1**	**8.5**	6.4	7.2	**9.3**	**9.9**	7.9	**9.9**	**11.4**
$c_1 = -7$, $c_2 = -2$	**12.7**	**10.9**	**8.5**	6.3	6.7	5.6	7.7	6.8	7.5	**9.8**	**11.9**
$c_1 = -5$, $c_2 = -5$	**10.2**	**9.0**	7.8	6.8	7.0	**8.7**	**8.1**	**8.5**	8.0	**9.1**	**11.1**
$c_1 = -7$, $c_2 = -5$	**11.0**	**9.5**	6.7	7.7	5.9	6.3	**8.4**	**8.1**	7.0	**8.9**	**11.3**
$c_1 = -7$, $c_2 = -7$	**10.1**	7.6	6.9	6.7	6.3	7.8	**8.2**	**8.1**	6.6	**8.4**	**11.0**

t(4) errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c_1 = c_2 = 0$	**21.8**	**16.0**	**14.2**	**10.5**	**11.6**	**9.6**	**11.3**	**12.4**	**12.3**	**15.2**	**19.2**
$c_1 = -2$, $c_2 = 0$	**15.4**	**12.5**	**10.2**	**8.1**	**8.6**	**9.7**	**9.6**	**9.2**	**10.3**	**12.8**	**16.4**
$c_1 = -5$, $c_2 = -0$	**15.1**	**12.3**	**8.9**	**8.8**	7.6	6.7	7.6	7.6	**9.8**	**12.0**	**14.1**
$c = -7$, $c_2 = -0$	**13.5**	**12.2**	**8.2**	6.7	5.3	6.1	6.7	7.5	**8.2**	**12.4**	**17.3**
$c_1 = -2$, $c_2 = -2$	**17.6**	**12.9**	**9.8**	**8.5**	7.8	7.4	**8.7**	**8.2**	**10.5**	**13.2**	**18.2**
$c_1 = -5$, $c_2 = -2$	**13.5**	**11.8**	**8.8**	6.0	6.4	**8.1**	7.8	**8.8**	**8.5**	**11.1**	**14.7**
$c_1 = -7$, $c_2 = -2$	**15.2**	**11.1**	6.7	7.9	6.5	6.0	6.8	7.1	**9.4**	**10.8**	**13.3**
$c_1 = -5$, $c_2 = -5$	**15.5**	**11.5**	**9.9**	7.2	5.7	5.6	6.0	7.9	**8.6**	**10.7**	**16.4**
$c_1 = -7$, $c_2 = -5$	**14.2**	**11.4**	**8.6**	5.9	6.0	6.7	7.5	**8.6**	**9.1**	**10.1**	**13.3**
$c_1 = -7$, $c_2 = -7$	**14.7**	**11.1**	**9.9**	7.3	4.7	5.1	5.8	7.7	7.5	**9.6**	**15.9**

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
$c_1 = c_2 = 0$	**21.1**	**16.3**	**13.0**	**11.4**	**8.7**	**9.0**	**9.4**	**10.6**	**12.3**	**14.4**	**19.5**
$c_1 = -2$, $c_2 = 0$	**20.9**	**12.4**	**12.5**	**9.1**	**8.1**	7.3	**8.6**	**8.6**	**11.0**	**15.7**	**18.5**
$c_1 = -5$, $c_2 = -0$	**17.1**	**12.9**	**9.3**	7.2	7.4	6.4	6.8	7.5	**9.8**	**13.9**	**17.9**
$c = -7$, $c_2 = -0$	**17.3**	**11.9**	**8.3**	7.3	4.5	5.9	5.6	6.9	**8.4**	**12.7**	**14.6**
$c_1 = -2$, $c_2 = -2$	**19.0**	**14.0**	**9.8**	**9.0**	6.8	6.1	7.8	**9.7**	**10.8**	**11.9**	**18.1**
$c_1 = -5$, $c_2 = -2$	**19.4**	**12.3**	**8.8**	6.9	7.5	5.5	6.8	6.6	**10.3**	**15.3**	**15.5**
$c_1 = -7$, $c_2 = -2$	**16.0**	**12.5**	**8.2**	7.0	6.0	6.2	6.4	7.5	**10.0**	**12.5**	**17.2**
$c_1 = -5$, $c_2 = -5$	**15.0**	**12.1**	6.6	7.4	6.3	5.8	5.3	**8.1**	**9.2**	**8.9**	**17.0**
$c_1 = -7$, $c_2 = -5$	**16.9**	**11.2**	**8.8**	6.8	6.3	5.1	5.8	5.3	**9.4**	**14.7**	**14.5**
$c_1 = -7$, $c_2 = -7$	**14.4**	**11.7**	6.3	6.7	6.4	4.9	4.6	7.3	**8.6**	**9.4**	**16.6**

Table 8: Size Performances(%) of IVX-QR with $\delta = 0.5$ ($n = 700$, $S = 1000$)

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
$c_1 = c_2 = 0$	**10.8**	**8.4**	8.0	7.8	6.4	6.5	7.0	6.3	8.0	6.1	**10.4**
$c_1 = -2$, $c_2 = 0$	**10.3**	7.9	6.6	3.9	5.2	5.0	5.6	5.4	5.6	6.8	7.6
$c_1 = -5$, $c_2 = -0$	**9.9**	**8.3**	6.5	4.8	4.5	5.2	5.4	4.2	5.3	6.6	**10.4**
$c = -7$, $c_2 = -0$	**9.8**	4.7	5.0	4.9	5.3	5.0	5.2	5.3	4.7	6.6	7.6
$c_1 = -2$, $c_2 = -2$	8.0	5.8	5.5	4.7	5.3	5.4	4.9	4.1	7.8	4.3	**8.5**
$c_1 = -5$, $c_2 = -2$	**8.6**	6.3	5.0	3.3	4.4	3.9	3.9	3.5	3.9	6.6	6.8
$c_1 = -7$, $c_2 = -2$	**8.7**	6.9	6.0	3.9	3.5	3.9	4.4	5.0	4.1	4.7	7.5
$c_1 = -5$, $c_2 = -5$	7.3	5.3	4.6	4.8	4.4	5.3	4.6	3.8	7.1	5.6	**8.4**
$c_1 = -7$, $c_2 = -5$	7.7	6.1	4.5	3.2	4.6	4.0	3.7	3.3	4.5	6.8	5.9
$c_1 = -7$, $c_2 = -7$	6.7	5.7	4.7	4.9	4.0	4.6	4.2	3.9	6.3	4.9	8.0

t(4) errors

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$c_1 = c_2 = 0$	**16.5**	**13.0**	8.0	6.9	6.0	5.2	6.7	8.0	6.4	**11.6**	**13.9**
$c_1 = -2$, $c_2 = 0$	**11.9**	**9.0**	7.5	5.9	4.8	5.1	4.6	5.6	6.7	**8.8**	**13.1**
$c_1 = -5$, $c_2 = -0$	**12.3**	8.0	6.9	4.9	5.3	3.5	5.3	5.1	6.3	**9.0**	**11.2**
$c = -7$, $c_2 = -0$	**10.9**	**11.4**	7.5	5.7	4.6	4.6	5.0	5.1	5.6	**8.3**	**11.8**
$c_1 = -2$, $c_2 = -2$	**13.7**	**10.1**	6.6	4.1	3.8	3.8	4.3	5.2	5.4	7.9	**11.5**
$c_1 = -5$, $c_2 = -2$	**10.3**	8.0	6.4	4.1	4.0	4.0	4.6	4.7	5.1	8.0	**9.6**
$c_1 = -7$, $c_2 = -2$	**10.7**	7.9	5.9	3.5	3.0	3.0	3.8	5.4	5.7	7.7	**10.7**
$c_1 = -5$, $c_2 = -5$	**11.7**	**8.9**	5.5	3.9	4.0	4.0	4.3	4.8	5.0	7.4	**9.5**
$c_1 = -7$, $c_2 = -5$	**9.5**	7.5	5.9	4.1	3.8	3.0	3.9	4.7	4.8	8.0	**9.5**
$c_1 = -7$, $c_2 = -7$	**10.8**	**8.5**	5.6	3.1	3.7	3.7	4.3	4.5	5.0	6.3	**9.2**

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
$c_1 = c_2 = 0$	**15.0**	**12.4**	7.9	7.4	6.2	7.4	6.3	7.6	7.1	**10.5**	**14.1**
$c_1 = -2$, $c_2 = 0$	**15.4**	**9.6**	8.8	4.3	6.3	4.4	4.3	7.0	7.3	**12.3**	**13**
$c_1 = -5$, $c_2 = -0$	**11.0**	**11.7**	6.6	6.5	4.1	4.5	4.1	6.7	7.2	**9.1**	**14.1**
$c = -7$, $c_2 = -0$	**10.9**	**10.7**	6.6	5.1	3.6	3.7	4.9	4.7	6.7	**9.7**	**9.7**
$c_1 = -2$, $c_2 = -2$	**11.5**	**9.3**	5.7	4.5	4.1	4.8	4.2	5.2	5.4	**9.0**	**9.2**
$c_1 = -5$, $c_2 = -2$	**12.7**	**8.8**	6.4	3.4	5.1	3.3	3.7	5.9	5.9	**9.8**	**10.5**
$c_1 = -7$, $c_2 = -2$	**11.7**	**10.6**	5.4	6.2	3.6	3.9	4.0	6.5	7.0	7.4	**11.8**
$c_1 = -5$, $c_2 = -5$	**11.4**	7.7	4.9	4.2	3.6	4.3	3.7	4.3	5.2	7.9	**8.5**
$c_1 = -7$, $c_2 = -5$	**10.6**	8.0	6.8	4.0	5.2	3.6	3.3	5.7	6.0	**10.5**	**9.5**
$c_1 = -7$, $c_2 = -7$	**11.4**	6.5	4.8	4.1	3.2	3.5	3.4	4.0	5.5	7.4	7.5

5 Quantile Predictability of Stock Returns

It is often standard practice in finance literature 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 briefly discuss the empirical rationale for predictive QR, which makes this approach attractive for stock return regressions. I then use the IVX-QR procedure to examine stock return predictability at various quantile levels with commonly used persistent predictors.

5.1 Empirical Motivation: Why Predictive QR?

To understand the empirical motivation for predictive QR, I begin with some selected stylized facts on financial assets. Cont (2001) notes that financial asset returns often have (i) conditional heavy tails, (ii) time-varying volatility, and (iii) asymmetric distributions. In stock return prediction tests, these stylized facts may motivate the use of QR, as now discussed.

To illustrate I consider a simple model for an asset return y_t:

$$y_t = \mu_{t-1} + \sigma_{t-1}\varepsilon_t \text{ with } \varepsilon_t \sim iid\ F_\varepsilon, \tag{5.1}$$

so that μ_{t-1} and σ_{t-1} signify the location (mean) and scale (volatility) conditional on $\mathcal{F}_{t-1}$. If we assume $E\left(\varepsilon_t|\mathcal{F}_{t-1}\right) = 0$, then $E\left(y_t|\mathcal{F}_{t-1}\right) = \mu_{t-1}$ and the predictive mean regression model uses $\mu_{t-1} = \beta_0 + \beta_1 x_{t-1}$ with predictor x_{t-1}. Inconclusive empirical results from mean regressions suggest that the magnitude of β_1 is small with most predictors x_{t-1}. Now also assume a stochastic volatility model: $\sigma_{t-1} = \theta_0 + \theta_1 x_{t-1}$. If a predictor x_{t-1} significantly influences the volatility of stock returns y_t, as many economic factors do, then $\theta_1 \neq 0$. In fact, the conditional variance often shows a greater systematic (or cyclical) variation than the conditional mean of stock returns, suggesting that $|\theta_1| > |\beta_1|$. The predictive QR now reduces to

$$Q_{y_t}\left(\tau|\mathcal{F}_{t-1}\right) = \beta_0 + \beta_1 x_{t-1} + \left(\theta_0 + \theta_1 x_{t-1}\right) Q_{\varepsilon_t}\left(\tau\right) = \beta_{0,\tau} + \beta_{1,\tau} x_{t-1}$$

where

$$\beta_{0,\tau} = \beta_0 + \theta_0 Q_{\varepsilon_t}\left(\tau\right) \text{ and } \beta_{1,\tau} = \beta_1 + \theta_1 Q_{\varepsilon_t}\left(\tau\right).$$

When ε_t is close to symmetric, $\beta_{1,\tau} \simeq \beta_1$ with $\tau = 0.5$. When β_1 is small, such as for a near-martingale stock return, $\beta_{1,\tau} \simeq \theta_1 Q_{\varepsilon_t}\left(\tau\right)$ and the absolute value of $\beta_{1,\tau}$ increases as $\tau \to 0$ (or $\tau \to 1$). This tendency may explain why the location (mean/median) of stock returns is more difficult to predict than other statistical measures, such as the scale (dispersion). Even when β_1 is not negligible, the magnitude of $\theta_1 Q_{\varepsilon_t}\left(\tau\right)$ would help in finding the predictability. Note also that when $\varepsilon_t \sim iid\ t\left(\nu\right)$ in (5.1), the absolute value of $\beta_{1,\tau}$ with $\tau \neq 0.5$ increases as ν decreases. The location of stock return predictability varies depending on the magnitude and sign of (β_1, θ_1), but QR can effectively locate it. Therefore, QR provides forecasts on stock return quantiles where the actual predictability is more likely to exist. This illustrates the rationale for using QR to predict stock returns with thick tails (smaller ν) and cyclical movements in conditional volatility (non-zero θ_1).

Lastly, consider an asymmetrically distributed ε_t in (5.1) with $|Q_{\varepsilon_t}(\tau)| > |Q_{\varepsilon_t}(1-\tau)|$ for a $\tau < 0.5$. In this case, predictor variables with nonneglible (β_1, θ_1) may predict lower quantiles of stock returns better than upper quantiles. These asymmetric predictable patterns and quantile-specific predictors can be detected through predictive QR.

Although the above example is stylized, we clearly see the advantages of QR in a stock return regression framework. Many of the results hypothesized above are confirmed in the empirical results below.

5.2 Empirical Results: IVX-QR Testing

I now show empirical results of stock return 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[6]. Following Cenesizoglu and Timmermann (2008), I classify the predictors into three categories.

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

[6] I would like to 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.

– net equity expansion (*ntis*)

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 various combinations of predictors. This application both complements and supplements the mean predictive regressions. There will be more applications of quantile-specific predictors that are useful for asset pricing and portfolio decisions-making.

The null test statistics in Proposition 3.2 is used with IVX filtering parameters $\delta = 0.5$ and $C_z = -5I_K$ because the specification showed uniformly good size properties for both single and multiple persistent predictors in the earlier section. Table 5 below reports the univariate regression results, where p-values (%) are rounded to one decimal place for exposition. The results shown in bold imply the rejection of the null hypothesis of no predictability at 5% level.

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	**1.0***	**0.1***	**0.3***	14.1	76.0	50.3	17.6	**0.5***	6.4	9.6	**1.8***
d/e	**0.0***	**0.0***	**0.0***	**0.4***	**1.8***	16.4	59.5	49.8	89.7	5.8	**0.3***
b/m	**0.2***	**0.0***	**0.1***	**0.4***	22.5	83.3	56.0	**2.8***	6.2	**0.8***	**1.7***
tbl	46.0	50.3	10.6	11.7	**3.6***	12.3	6.0	19.9	**3.9***	**0.6***	**0.3***
dfy	33.6	66.3	98.5	89.8	53.5	5.6	**3.0***	**0.1***	**0.0***	**0.0***	**0.1***
e/p	75.3	89.8	42.1	51.4	96.5	71.2	89.7	47.2	55.7	33.0	32.6
ntis	23.0	96.7	24.1	13.2	10.1	73.8	71.6	74.1	65.9	94.8	32.5
tms	41.8	15.6	82.5	68.2	26.6	36.7	14.6	58.1	23.2	13.0	16.3

The result is roughly consistent with the results of Maynard et al. (2011) and Cenesizoglu and Timmermann (2008). I find significant upper quantile predictive ability for the *tbl* and *dfy*, and lower quantile predictive power for the *d/e*. Evidence of both lower and upper quantile predictability from *b/m* is also similar. One notable difference is the evidence of predictability at lower quantiles with *d/p*. Overall, I find little evidence of predictability at the median. The results confirm many hypothesized empirical results in the earlier section - the weak predictability at the mean/median of stock returns, the stronger forecasting capability at quantiles away from the median and several quantile specific (lower, upper or both quantiles) predictors .

For multivariate regression applications, I use selective predictor combinations for illustrative purposes. The selection scheme is as follows: First, I choose significant predictors from univariate regression results (*d/p, d/e, b/m, tbl,* and *dfy* in this instance). Second, I classify the chosen predictors into three groups - group L, group B and group U. Group L corresponds to the group with significant lower-quantile predictors, and so on for the other groups. I select one predictor from each group to produce a bivariate predictor.

- Group of lower quantile predictors (Group L): *d/p* and *d/e*
- Both upper and lower quantile predictor (Group B): *b/m*
- Group of upper quantile predictors (Group U): *tbl* and *dfy*

Finally, I choose predictor combinations exhibiting little evidence of comovement between the predictors. Although evidence of comovement between predictors does not completely reduce the appeal of the combinations; we may prefer less-comoving systems for better forecast models[7].

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 Appendix 7.6). I find little evidence of comovement between (*d/p, tbl*), (*d/e, dfy*), (*d/e, b/m*), (*tbl, bm*) and (*dfy, b/m*).

The above selection scheme is used primarily for illustrative purposes, and I do not rule out the possibility of significant results from other combinations[8]. However, it is partly justifiable from a theoretical perspective. For example, both *d/p* and *b/m* are ratios measuring undervaluation in the stock market and are thus positively correlated to subsequent returns, while *tbl* is a macro variable that may have different predictive patterns. If we choose between (*d/p, b/m*) and (*tbl, b/m*), the above economic rational recommends (*tbl, b/m*) because it shares fewer common characteristics. Diagnostic tests indicate strong comoving evidence between *d/p* and *b/m* but not between *tbl* and *b/m*, corresponding with the economic foundation. Similar arguments hold for the group of corporate finance variables measuring managerial financing activity. Managers' timing of issuing equity precedes falling stock prices, indicating negative correlation with stock returns.

From Table 6 below, I confirm that many predictors are jointly significant at various quantiles with much stronger evidences than univariate regressions. Many existing studies only considered a single persistent predictor when they considered the spurious predictability from the size distortion. 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.

[7] 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 nonsingular Ω_{xx} in (2.10)) will be left for future research.

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

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

Multivariate regressions with two predictors among *d/p, d/e, tbl, dfy* and *b/m*

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p, tbl	**0.4***	**0.1***	**0.8***	**2.4***	**1.0***	7.6	8.6	**0.0***	**1.2***	**1.0***	**0.0***
d/e, tbl	**0.0***	**0.0***	**0.0***	**0.2***	**0.1***	**0.7***	6.8	36.6	6.0	**0.0***	**0.0***
d/e, dfy	13.4	3.6*	13.0	20.2	33.0	11.0	**2.6***	**0.2***	**0.0***	**0.5***	**0.2***
d/e, b/m	**0.0***	**0.0***	**0.0***	**0.1***	**1.2***	57.2	40.6	16.3	**2.1***	**0.0***	**0.0***
tbl, b/m	**0.6***	**0.0***	**0.2***	**2.7***	6.0	**4.6***	**0.6***	**0.6***	**0.1***	**0.2***	**2.9***
dfy, b/m	**0.9***	**0.2***	**0.9***	9.4	12.5	14.6	19.1	**0.1***	**0.1***	**0.0***	**0.3***

I run the same analysis 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 change or improved market efficiency. Table 7 below shows much weaker predictability evidences, but meaningful 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 various quantiles, which might explain the conflicts between the mean regression results and provide empirical evidence for a resolution to the debate.

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, 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	75.6	73.0	57.8	30.4	8.1	84.6	**1.4***	**0.1***	15.7	76.1	19.8
d/e	75.5	93.1	75.9	69.8	64.6	49.3	80.3	22.8	92.6	50.2	57.8
b/m	**0.1***	**0.0***	7.0	**2.4***	7.1	48.7	68.5	18.6	33.2	20.6	19.0
tbl	37.8	7.3	**0.1***	**0.3***	**0.1***	**3.1***	**1.1***	6.9	5.4	**0.5***	**1.1***
dfy	14.6	32.5	76.7	62.1	16.8	8.6	13.9	**1.8***	**0.2***	**3.1***	15.1
e/p	59.2	82.0	70.7	51.8	40.9	**2.5***	23.6	**1.4***	98.8	80.8	6.6
ntis	57.6	8.2	28.6	43.5	52.2	12.1	9.3	22.1	18.5	21.3	8.0
tms	67.5	88.8	26.3	67.0	25.5	10.1	28.0	22.6	30.1	15.9	5.7

I proceed to models with two predictors using the earlier selection scheme. From the same diagnostic tests, I find little evidence of comovement between (b/m, tbl) and (tbl, dfy). From Table 8 below, we see new empirical support for stock return forecast models for post-1952 periods, using (*b/m, tbl*) or (*tbl, dfy*). It turns out that the combination of one valuation ratio (*b/m*) and a macro variable (*tbl*), or the latter with a measure of default risk (*dfy*), provide a potentially improved forecast model for stock returns. IVX-QR corrections ensure that the predictability results are not spurious despite the multiple persistent predictors. We may continue to the model with these three factors.

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

Multivariate regressions with two predictors among *b/m, d/p, dfy, e/p* and *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***	**0.0***	5.2	**1.1***	**0.6***	**2.2***	**1.1***	**1.5***	**1.3***	**0.0***	**2.6***
dfy, tbl	17.1	36.3	**3.8***	**0.5***	**1.2***	**0.4***	**0.2***	**0.8***	**1.6***	**2.6***	7.5

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 disappears at the median. A partial answer to the empirical puzzle of stock return mean/median predictability may be provided by the results. 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, the forecasting capability at most quantiles can be substantially enhanced relative to a model with a single predictor. The predictive ability 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 (type I error) arising from multiple manifestations of predictor persistence. 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 where the evidence for multiple significant predictors with uncertain degrees of persistence is overwhelming.

In our 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. Stock return forecast models based on quantile-specific predictors may be used for portfolio optimization problem of a risk-averse investor, who considers the conditional distribution of future excess returns based on some information set in the current period. This application corresponds to a common practice in financial economics, see, e.g., Cenesizoglu and Timmermann (2008, Section 5.1). In risk management applications, some significant lower quantile predictors of stock returns play a role in estimating the expected loss at a given probability (quantile) level. Finding significant stock return quantile predictors among potential candidates precedes all these practical applications. 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 at most quantiles suggests there is scope for further improvement in a wider class of 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. Further guidance on the determination of the IVX persistence choice parameter (δ) is important. It is possible, for instance, to use (partial) information on predictor persistence and quantile specific characteristics such as the degree of tail dependence in this selection. Finally, with regard to the asymptotics, improved IVX-QR inference at extreme quantiles, especially under strong persistence and tail dependence, may be possible using extremal QR limit theory.

7 Appendix

Some proofs directly come from the existing papers, such as Magdalinos and Phillips (2012a,b; MP_a and MP_b, respectively), and Phillips and Lee (2012b: PL_b).

7.1 Proofs for Section 2.2

The following lemma provides the asymptotics of the processes driving the limit theory of $\hat{\beta}_{\tau}^{QR}$ in (2.11).

Lemma 7.1 *1.*

$$G_{\tau,n}^{x} := D_n^{-1} \sum_{t=1}^{n} X_{t-1} \psi_{\tau}\left(u_{0t\tau}\right) \Longrightarrow G_{\tau}^{x},$$

where

$$G_{\tau}^{x} = \begin{cases} N\left(0, \tau(1-\tau)\begin{bmatrix} 1 & 0 \\ 0 & \Omega_{xx} \end{bmatrix}\right) & \text{for (I0)}, \\ N\left(0, \tau(1-\tau)\begin{bmatrix} 1 & 0 \\ 0 & V_{xx} \end{bmatrix}\right) & \text{for (MI)}, \\ \begin{bmatrix} B_{\psi}(1) \\ \int J_x^c(r) dB_{\psi} \end{bmatrix} & \text{for (I1)}, \\ MN\left[0, \tau(1-\tau)\begin{bmatrix} 1 & 0 \\ 0 & \tilde{V}_{xx} \end{bmatrix}\right], & \text{for (ME)}. \end{cases}$$

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* $Y_C \equiv N\left(0, \int_0^{\infty} e^{-rC} \Omega_{xx} e^{-rC} dr\right)$.

2.

$$M_{\beta_{\tau},n}^{x} := D_n^{-1} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}\left(0\right) X_{t-1} X_{t-1}' D_n^{-1} \Longrightarrow M_{\beta_{\tau}}^{x}.$$

where

$$M_{\beta_{\tau}}^{x} = \begin{cases} f_{u_{0\tau}}\left(0\right)\begin{bmatrix} 1 & 0 \\ 0 & \Omega_{xx} \end{bmatrix} & \text{for (I0)}, \\ f_{u_{0\tau}}\left(0\right)\begin{bmatrix} 1 & 0 \\ 0 & V_{xx} \end{bmatrix} & \text{for (MI)}, \\ f_{u_{0\tau}}\left(0\right)\begin{bmatrix} 1 & \int J_x^c(r)' \\ \int J_x^c(r) & \int J_x^c(r) J_x^c(r)' \end{bmatrix} & \text{for (I1)}, \\ f_{u_{0\tau}}\left(0\right)\begin{bmatrix} 1 & 0 \\ 0 & \tilde{V}_{xx} \end{bmatrix} & \text{for (ME)}. \end{cases}$$

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

Proof. All proofs for (I0) cases are standard so omitted, see e.g., Koenker (2005, Theorem 4.1).

1. For (I1) case, from (2.10),

$$G^x_{\tau,n} = D_n^{-1} \sum_{t=1}^{n} X_{t-1}\psi_\tau\left(u_{0t\tau}\right) = \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_\tau\left(u_{0t\tau}\right) \\ \frac{1}{n} \sum_{t=1}^{n} x_{t-1}\psi_\tau\left(u_{0t\tau}\right) \end{bmatrix} \Longrightarrow \begin{bmatrix} B_\psi(1) \\ \int J^c_x(r) dB_\psi \end{bmatrix}.$$

For (MI) case,

$$D_n^{-1} \sum_{t=1}^{n} X_{t-1}\psi_\tau\left(u_{0t\tau}\right) = \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_\tau\left(u_{0t\tau}\right) \\ \frac{1}{n^{\frac{1+\alpha}{2}}} \sum_{t=1}^{n} x_{t-1}\psi_\tau\left(u_{0t\tau}\right) \end{bmatrix} \Longrightarrow N\left(0, \tau(1-\tau)\begin{bmatrix} 1 & 0 \\ 0 & V_{xx} \end{bmatrix}\right),$$

where $V_{xx} := \int_0^\infty e^{rC}\Omega_{xx}e^{rC}dr$ since

$$\begin{aligned} D_n^{-1} \sum_{t=1}^{n} E\left[\psi_\tau\left(u_{0t\tau}\right)^2 X_{t-1}X'_{t-1}|\mathcal{F}_{t-1}\right] D_n^{-1} &= \begin{bmatrix} \tau(1-\tau) & \frac{\tau(1-\tau)}{n^{1+\frac{\alpha}{2}}} \sum_{t=1}^{n} x'_{t-1} \\ \frac{\tau(1-\tau)}{n^{1+\frac{\alpha}{2}}} \sum_{t=1}^{n} x_{t-1} & \frac{\tau(1-\tau)}{n^{1+\alpha}} \sum_{t=1}^{n} x_{t-1}x'_{t-1} \end{bmatrix} \\ &\to {}^p \tau(1-\tau)\begin{bmatrix} 1 & 0 \\ 0 & V_{xx} \end{bmatrix}, \end{aligned}$$

using the result of MP_a :

$$\frac{1}{n^{\frac{1}{2}+\alpha}} \sum_{t=1}^{n} x'_{t-1} = O_p(1) \text{ and } \frac{1}{n^{1+\alpha}} \sum_{t=1}^{n} x_{t-1}x'_{t-1} \to^p V_{xx}.$$

Therefore, the stability condition for MGCLT is established. The conditional Lindeberg condition is easy to show since $\psi_\tau\left(u_{0t\tau}\right)$ is mds, so is omitted.

Now for (ME) case,

$$D_n^{-1} \sum_{t=1}^{n} X_{t-1}\psi_\tau\left(u_{0t\tau}\right) = \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_\tau\left(u_{0t\tau}\right) \\ \frac{1}{n^\alpha} R_n^{-n} \sum_{t=1}^{n} x_{t-1}\psi_\tau\left(u_{0t\tau}\right) \end{bmatrix} \Longrightarrow MN\left[0, \tau(1-\tau)\begin{bmatrix} 1 & 0 \\ 0 & \tilde{V}_{xx} \end{bmatrix}\right].$$

where $\tilde{V}_{xx} := \int_0^\infty e^{-rC}Y_C Y'_C e^{-rC}dr$ and $Y_C \equiv N\left(0, \int_0^\infty e^{-rC}\Omega_{xx}e^{-rC}dr\right)$ as in MP_a. The proof is similar to (MI) case so is omitted.

2. For (I1) case,

$$\begin{aligned} M^x_{\beta_\tau,n} &= D_n^{-1} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, X_{t-1}X'_{t-1}D_n^{-1} \\ &= \begin{bmatrix} \frac{1}{n} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0) & \frac{1}{n^{\frac{3}{2}}} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, x'_{t-1} \\ \frac{1}{n^{\frac{3}{2}}} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, x_{t-1} & \frac{1}{n^2} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, \tilde{z}_{t-1}\tilde{z}'_{t-1} \end{bmatrix} \\ &\Longrightarrow \begin{bmatrix} f_{u_{0\tau}}(0) & f_{u_{0\tau}}(0) \int J^c_x(r)' \\ f_{u_{0\tau}}(0) \int J^c_x(r) & f_{u_{0\tau}}(0) \int J^c_x(r) J^c_x(r)' \end{bmatrix}, \end{aligned}$$

where, from the standard FCLT and CMT (Phillips, 1987),

$$
\begin{aligned}
\frac{1}{n^{\frac{3}{2}}}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, x'_{t-1} &= \frac{1}{n^{\frac{3}{2}}}\sum_{t=1}^{n}\left(f_{u_{0t\tau},t-1}(0) - f_{u_{0\tau}}(0)\right) x'_{t-1} + f_{u_{0\tau}}(0)\,\frac{1}{n^{\frac{3}{2}}}\sum_{t=1}^{n} x'_{t-1} \\
&= f_{u_{0\tau}}(0)\sum_{t=1}^{n}\left(\frac{x'_{t-1}}{\sqrt{n}}\right)\frac{1}{n} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\left(f_{u_{0t\tau},t-1}(0) - f_{u_{0\tau}}(0)\right)}{\sqrt{n}}\left(\frac{x_{t-1}}{\sqrt{n}}\right) \\
&= f_{u_{0\tau}}(0)\sum_{t=1}^{n}\left(\frac{x'_{t-1}}{\sqrt{n}}\right)\frac{1}{n} + O_p\left(\frac{1}{\sqrt{n}}\right) \Longrightarrow f_{u_{0\tau}}(0)\int J_x^c(r)'.
\end{aligned}
$$

Other elements can be shown similarly, hence omitted.

For (MI) case, using the same method and the fact $n^{-1/2-\alpha}\sum_{t=1}^{n} x'_{t-1} = O_p(1)$ again,

$$
\begin{aligned}
M^x_{\beta_\tau,n} &= D_n^{-1}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, X_{t-1}X'_{t-1}D_n^{-1} \\
&= \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0) & \frac{1}{n^{1+\frac{\alpha}{2}}}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, x'_{t-1} \\ \frac{1}{n^{1+\frac{\alpha}{2}}}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, x_{t-1} & \frac{1}{n^{1+\alpha}}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, x_{t-1}x'_{t-1} \end{bmatrix} \to^p \begin{bmatrix} f_{u_{0\tau}}(0) & 0 \\ 0 & f_{u_{0\tau}}(0)\, V_{xx} \end{bmatrix}.
\end{aligned}
$$

(ME) case will be shown in the exact same way,

$$
\begin{aligned}
M^x_{\beta_\tau,n} &= D_n^{-1}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, X_{t-1}X'_{t-1}D_n^{-1} \\
&= \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0) & \frac{1}{n^{\frac{1}{2}+\alpha}}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, x'_{t-1}R_n^{-n} \\ \frac{1}{n^{\frac{1}{2}+\alpha}}R_n^{-n}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, x_{t-1} & \frac{1}{n^{2\alpha}}R_n^{-n}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, x_{t-1}x'_{t-1}R_n^{-n} \end{bmatrix} \\
&\Longrightarrow \begin{bmatrix} f_{u_{0\tau}}(0) & 0 \\ 0 & f_{u_{0\tau}}(0)\,\tilde{V}_{xx} \end{bmatrix}.
\end{aligned}
$$

■

Proof of Theorem 2.1. As in Xiao (2009, Proof of Theorem 1), we can linearize (2.11) in terms of an arbitrary centred quantity $D_n^{-1}\left(\hat{\beta}_\tau - \beta_\tau\right)$ using Knight's identity (Knight, 1989). Note that (2.11) is a convex minimization problem. Using the convexity lemma (Pollard, 1991) we can take the distributional limit (or probability limit in the expanded probability space) of the linearized (2.11) first, and then minimize to get $D_n^{-1}\left(\hat{\beta}_\tau^{QR} - \beta_\tau\right)$. It directly leads to:

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

and the results of Theorem 2.1 follow from Lemma 7.1. ■

7.2 Dequantiling Problem

I show the dequantiled dependent variable in (3.1) asymptotically does not cause any problem in the IVX-QR implementation. This is a special feature of IVX instrumentation. To see this, first consider the ordinary nonstationary quantile regression, if we use the pre-estimated $\sqrt{n}$-consistent estimator $\widehat{\beta}_{0,\tau}$,

$$\begin{aligned} & \frac{1}{n}\sum_{t=1}^{n} x_{t-1}\psi_\tau\left(y_t-\beta_{0,\tau}-\beta_{1,\tau}x_{t-1}\right)-\frac{1}{n}\sum_{t=1}^{n} x_{t-1}\psi_\tau\left(y_t-\widehat{\beta}_{0,\tau}-\beta_{1,\tau}x_{t-1}\right) \\ = & \frac{1}{n}\sum_{t=1}^{n} x_{t-1}\left\{1\left(u_{0t\tau}\leq\widehat{\beta}_{0,\tau}-\beta_{0,\tau}\right)-1\left(u_{0t\tau}\leq 0\right)\right\}, \end{aligned}$$

and by taking conditional expectation $E_{t-1}(\cdot)$:

$$\begin{aligned} = & \frac{1}{n}\sum_{t=1}^{n} x_{t-1}E_{t-1}\left\{1\left(u_{0t\tau}\leq\widehat{\beta}_{0,\tau}-\beta_{0,\tau}\right)-1\left(u_{0t\tau}\leq 0\right)\right\} \\ \sim & f_{u_\tau}(0)\frac{1}{n}\sum_{t=1}^{n} x_{t-1}\left(\widehat{\beta}_{0,\tau}-\beta_{0,\tau}\right)=O\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} x_{t-1}\right)=O_p(1), \end{aligned}$$

hence the effect of replacing $\beta_{0,\tau}$ by $\widehat{\beta}_{0,\tau}$ is still present in the limit.

On the other hand, in the IVX-QR procedure, from (3.7):

$$\begin{aligned} & \sum_{t=1}^{n} \widetilde{Z}_{t-1}\psi_\tau\left(y_t-\beta_{0,\tau}-\beta_{1,\tau}x_{t-1}\right)-\sum_{t=1}^{n} \widetilde{Z}_{t-1}\psi_\tau\left(y_t-\widehat{\beta}_{0,\tau}-\beta_{1,\tau}x_{t-1}\right) \\ = & \sum_{t=1}^{n} \widetilde{Z}_{t-1}\left\{1\left(u_{0t\tau}\leq\widehat{\beta}_{0,\tau}-\beta_{0,\tau}\right)-1\left(u_{0t\tau}\leq 0\right)\right\}. \end{aligned} \tag{7.1}$$

Assume $\alpha=1$ for simplicity, since other cases will be similar. By taking conditional expectation $E_{t-1}(\cdot)$:

$$\begin{aligned} = & \frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n} \tilde{z}_{t-1}E_{t-1}\left\{1\left(u_{0t\tau}\leq\widehat{\beta}_{0,\tau}-\beta_{0,\tau}\right)-1\left(u_{0t\tau}\leq 0\right)\right\} \\ \sim & f_{u_\tau}(0)\frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n} \tilde{z}_{t-1}\left(\widehat{\beta}_{0,\tau}-\beta_{0,\tau}\right)=O\left(\frac{1}{n^{1+\frac{\delta}{2}}}\sum_{t=1}^{n} \tilde{z}_{t-1}\right)=o_p(1), \end{aligned}$$

since $\sum_{t=1}^{n}\tilde{z}_{t-1}=O_p\left(n^{\frac{1}{2}+\delta}\right)$ from the signal strength of mildly integrated process (MP_a), so $n^{-(1+\delta/2)}\sum_{t=1}^{n}\tilde{z}_{t-1}=O_p\left(n^{-(1-\delta)/2}\right)=o_p(1)$. Now it is easy to show the predictable quadratic variation is degenerate, i.e.,

$$\frac{1}{n^{1+\delta}}\sum_{t=1}^{n} E_{t-1}\left(\tilde{z}_{t-1}\tilde{z}_{t-1}'\zeta_t^2\right)\leq\left(\frac{1}{n^{1+\delta}}\sum_{t=1}^{n}\tilde{z}_{t-1}\tilde{z}_{t-1}'\right)\sup E_{t-1}\left(\zeta_t^2\right)=o_p(1),$$

where

$$\zeta_t = 1\left(u_{0t\tau} \le \widehat{\beta}_{0,\tau} - \beta_{0,\tau}\right) - 1\left(u_{0t\tau} \le 0\right)$$

using the consistency of $\widehat{\beta}_{0,\tau}$. Therefore (7.1) is $o_p(1)$ and using the pre-estimated $\sqrt{n}$-consistent estimator $\widehat{\beta}_{0,\tau}$ does not have any problem in the IVX-QR implementation.

7.3 Intercept Inference

Assume $\alpha = 1$ for simplicity. Define

$$\hat{\beta}_{0,\tau} = \arg\min_{\beta_0} \sum \rho_\tau\left(y_t - \hat{\beta}_{1,\tau}^{IVX-QR\prime} x_{t-1} - \beta_0\right),$$

then it is easy to show

$$\left(\hat{\beta}_{0,\tau}^{IVX-QR} - \beta_{0,\tau}\right) = \left[\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\right]^{-1} \left\{\sum_{t=1}^{n} \psi_t(u_{0t\tau}) + \left(\hat{\beta}_{1,\tau}^{IVX-QR} - \beta_{1,\tau}\right) \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, x_{t-1}\right\}$$

hence

$$\begin{aligned} n^{\frac{\delta}{2}}\left(\hat{\beta}_{0,\tau}^{IVX-QR} - \beta_{0,\tau}\right) &= n^{\frac{1+\delta}{2}}\left(\hat{\beta}_{1,\tau}^{IVX-QR} - \beta_{1,\tau}\right) f_{u_\tau}(0) \left(\frac{1}{n^{3/2}}\sum_{t=1}^{n} x_{t-1}\right)\left[\frac{1}{n}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\right]^{-1} + o_p(1) \\ &\sim n^{\frac{1+\delta}{2}}\left(\hat{\beta}_{1,\tau}^{IVX-QR} - \beta_{1,\tau}\right)\int J_x^c. \end{aligned}$$

Therefore, joint mixed normal limit theory for $\left\{n^{\frac{\delta}{2}}\left(\hat{\beta}_{0,\tau}^{IVX-QR} - \beta_{0,\tau}\right), n^{\frac{1+\delta}{2}}\left(\hat{\beta}_{1,\tau}^{IVX-QR} - \beta_{1,\tau}\right)\right\}$ is available and inference is possible through the self-normalization.

7.4 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 $\hat{\beta}_{1,\tau}^{IVXQR}$ is identical to that of the ordinary QR estimator $\hat{\beta}_{1,\tau}$ in Theorem 2.1. For (I0), (MI) and (I1) cases in general, $\tilde{z}_{t-1}$ reduces the persistence of x_{t-1} when x_{t-1} is more persistent than $\tilde{z}_{t-1}$ ($\delta < \alpha$). If x_{t-1} is less persistent than $\tilde{z}_{t-1}$ ($\delta > \alpha$) or x_{t-1} is (I0), then original persistence of x_{t-1} is maintained, hence the ordinary QR limit theory is achieved. When x_{t-1} is (ME), the limit theory is somewhat different. The remainder term of IVX dominates the asymptotics, a point on which Phillips and Lee (2012b) provided details in 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.*

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

2.

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

3.

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

Proof. 1. (i) For (MI) and (I1) cases with $0 < \delta < \min(\alpha, 1)$, from MP$_b$ (Proposition A2 and lemma 3.1)

$$\begin{aligned} \frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n}\tilde{z}_{t-1}\psi_\tau\left(u_{0t\tau}\right) &= \frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n} z_{t-1}\psi_\tau\left(u_{0t\tau}\right) + \frac{C}{n^{\frac{1+\delta}{2}+\alpha}}\sum_{t=1}^{n}\eta_{nt-1}\psi_\tau\left(u_{0t\tau}\right) \\ &= \frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n} z_{t-1}\psi_\tau\left(u_{0t\tau}\right) + o_p(1) \end{aligned}$$

then $G_{\tau,n} = \sum_{t=1}^{n}\xi_{nt} + o_p(1)$ where

$$\sum_{t=1}^{n}\xi_{nt} := \sum_{t=1}^{n}\frac{1}{n^{\frac{1+\delta}{2}}} z_{t-1}\psi_\tau\left(u_{0t\tau}\right).$$

The Stability condition for MGCLT is easy to show:

$$\sum_{t=1}^{n} E\left[\xi_{nt}\xi'_{nt}|\mathcal{F}_{t-1}\right] = \tau\left(1-\tau\right)\frac{1}{n^{1+\delta}}\sum_{t=1}^{n} z_{t-1}z'_{t-1} \to^p \tau(1-\tau)V^{x}_{zz}.$$

For any $\epsilon > 0$

$$\begin{aligned}
& \sum_{t=1}^{n} E\left[\|\xi_{nt}\|^2 1\left(\|\xi_{nt}\| > \epsilon\right) |\mathcal{F}_{t-1}\right] \\
= & \sum_{t=1}^{n} E\left[\|\xi_{nt}\|^2 1\left(\|z_{t-1}\psi_\tau\left(u_{0t\tau}\right)\| > \epsilon n^{\frac{1+\delta}{2}}\right) |\mathcal{F}_{t-1}\right] \\
\leq & \sum_{t=1}^{n} E\left[\|\xi_{nt}\|^2 1\left(|\psi_\tau\left(u_{0t\tau}\right)| > \epsilon n^{\frac{\delta}{4}}\right) |\mathcal{F}_{t-1}\right] + \sum_{t=1}^{n} E\left[\|\xi_{nt}\|^2 1\left(\|z_{t-1}\| > \epsilon n^{\frac{1}{2}+\frac{\delta}{4}}\right) |\mathcal{F}_{t-1}\right] \\
= & \left(\frac{1}{n^{1+\delta}} \sum_{t=1}^{n} \|z_{t-1}\|^2\right) E\left[|\psi_\tau\left(u_{0t\tau}\right)|^2 1\left(|\psi_\tau\left(u_{0t\tau}\right)|^2 > \epsilon^2 n^{\frac{\delta}{2}}\right) |\mathcal{F}_{t-1}\right] \\
& + \left(\frac{1}{n^{1+\delta}} \sum_{t=1}^{n} \|z_{t-1}\|^2 1\left(\|z_{t-1}\|^2 > \epsilon^2 n^{1+\frac{\delta}{2}}\right)\right) E\left[|\psi_\tau\left(u_{0t\tau}\right)|^2 |\mathcal{F}_{t-1}\right] \\
= & o_p(1) + \left(\frac{1}{n} \sum_{t=1}^{n} \left\|\frac{z_{t-1}}{n^{\frac{\delta}{2}}}\right\|^2 1\left(\left\|\frac{z_{t-1}}{n^{\frac{\delta}{2}}}\right\|^2 > \epsilon n^{1-\frac{\delta}{2}}\right)\right) \tau(1-\tau) \\
= & o_p(1),
\end{aligned}$$

confirming the conditional Lindeberg condition. Thus, from MGCLT (Hall and Heyde, 1980):

$$G_{\tau,n} \Longrightarrow N\left(0, \tau(1-\tau)V_{zz}^x\right).$$

(ii) For (MI) and (I1) cases with $\alpha \in (0, \delta)$, again from MP_b (Proposition A2 and lemma 3.5)

$$\frac{1}{n^{\frac{1+\alpha}{2}}} \sum_{t=1}^{n} \tilde{z}_{t-1}\psi_\tau\left(u_{0t\tau}\right) = \frac{1}{n^{\frac{1+\alpha}{2}}} \sum_{t=1}^{n} x_{t-1}\psi_\tau\left(u_{0t\tau}\right) + o_p(1),$$

and by the similar procedure above,

$$G_{\tau,n} = \sum_{t=1}^{n} \frac{1}{n^{\frac{1+\alpha}{2}}} x_{t-1}\psi_\tau\left(u_{0t\tau}\right) \Longrightarrow N\left(0, \tau(1-\tau)V_{xx}\right).$$

(iii) If x_{t-1} belongs to (I0), using Lemma A.2 in Kostakis et al. (2012), it is easy to show that

$$G_{\tau,n} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \tilde{z}_{t-1}\psi_\tau\left(u_{0t\tau}\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_{t-1}\psi_\tau\left(u_{0t\tau}\right) + o_p(1) \Longrightarrow N\left(0, \tau(1-\tau)\Omega_{xx}\right).$$

2. (i) For (MI) and (I1) cases with $0 < \delta < \min(\alpha, 1)$, note that

$$\begin{aligned}
M_{\gamma_\tau,n} &= \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\,\tilde{Z}_{t-1,n}\tilde{Z}'_{t-1,n} = \frac{1}{n^{1+\delta}}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\,\tilde{z}_{t-1}\tilde{z}'_{t-1} \\
&= \frac{1}{n^{1+\delta}}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\,z_{t-1}z'_{t-1} + o_p(1) \text{ (from MP}_b\text{, lemma 3.1-(iii))} \\
&= f_{u_{0\tau}}(0)\left(\frac{1}{n^{1+\delta}}\sum_{t=1}^{n} z_{t-1}z'_{t-1}\right) + \frac{1}{n^{1+\delta}}\sum_{t=1}^{n}\left(f_{u_{0t\tau},t-1}(0) - f_{u_{0\tau}}(0)\right)z_{t-1}z'_{t-1} + o_p(1) \\
&\to^{p} f_{u_{0\tau}}(0)\,V_{zz}^{x}
\end{aligned}$$

where we used Assumption 2.1-(i) and the result $\sup_t\left(\frac{z_{t-1}}{n^{\delta/2}}\right) = O_p(1)$ from MP_a so that

$$\begin{aligned}
&\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n}\left(f_{u_{0t\tau},t-1}(0) - f_{u_{0\tau}}(0)\right)z_{t-1}z'_{t-1} \\
&= \sum_{t=1}^{n}\left(\frac{f_{u_{0t\tau},t-1}(0) - f_{u_{0\tau}}(0)}{\sqrt{n}}\right)\left(\frac{z_{t-1}}{n^{\delta/2}}\right)\left(\frac{z_{t-1}}{n^{\delta/2}}\right)' = O_p(1),
\end{aligned}$$

hence

$$\frac{1}{n^{1+\delta}}\sum_{t=1}^{n}\left(f_{u_{0t\tau},t-1}(0) - f_{u_{0\tau}}(0)\right)z_{t-1}z'_{t-1} = O_p(\frac{1}{\sqrt{n}}) = o_p(1).$$

(ii) For (MI) and (I1) cases with $\alpha \in (0, \delta)$,

$$\begin{aligned}
M_{\gamma_\tau,n} &= \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\,\tilde{Z}_{t-1,n}\tilde{Z}'_{t-1,n} = \frac{1}{n^{1+\alpha}}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\,\tilde{z}_{t-1}\tilde{z}'_{t-1} \\
&= \frac{1}{n^{1+\alpha}}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\,x_{t-1}x'_{t-1} + o_p(1) \text{ (from MP}_b\text{, lemma 3.5-(ii))} \\
&= f_{u_{0\tau}}(0)\left(\frac{1}{n^{1+\alpha}}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\,x_{t-1}x'_{t-1}\right) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{f_{u_{0t\tau},t-1}(0) - f_{u_{0\tau}}(0)}{\sqrt{n}}\right)\left(\frac{x_{t-1}}{n^{\alpha/2}}\right)\left(\frac{x_{t-1}}{n^{\alpha/2}}\right)' + o_p(1) \\
&\to^{p} f_{u_{0\tau}}(0)\,V_{xx}
\end{aligned}$$

(iii) If x_{t-1} belongs to (I0), using Lemma A.2 in Kostakis et al. (2012) again,

$$M_{\gamma_\tau,n} = \frac{1}{n}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\,\tilde{z}_{t-1}\tilde{z}'_{t-1} = \frac{1}{n}\sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\,x_{t-1}x'_{t-1} + o_p(1) \to^{p} f_{u_{0\tau}}(0)\,\Omega_{xx}$$

3. (i) For (MI) and (I1) cases with $0 < \delta < \min(\alpha, 1)$,

$$
\begin{aligned}
M_{\beta_\tau,n} &= \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, \tilde{Z}_{t-1,n} X'_{t-1,n} \\
&= \frac{1}{n^{1+\delta}} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, \tilde{z}_{t-1} x'_{t-1} = f_{u_{0\tau}}(0) \left\{ \frac{1}{n^{1+\delta}} \sum_{t=1}^{n} \tilde{z}_{t-1} x'_{t-1} \right\} + o_p(1) \text{ as earlier,} \\
&= f_{u_{0\tau}}(0) \left\{ \frac{1}{n^{1+\delta}} \sum_{t=1}^{n} z_{t-1} x'_{t-1} - \frac{C_z^{-1} C}{n^{1+\alpha}} \sum_{t=1}^{n} x_{t-1} x'_{t-1} \right\} + o_p(1) \text{ (from MP}_b\text{, lemma 3.1-(ii))} \\
&\Longrightarrow \begin{cases} f_{u_{0\tau}}(0) \left\{ -C_z^{-1} \left(\int dB_x J_x^{c\prime} + \Omega_{xx} \right) - C_z^{-1} C \left(\int J_x^c J_x^{c\prime} \right) \right\}, & \text{if } \alpha = 1 \\ f_{u_{0\tau}}(0) \left\{ -C_z^{-1} \Omega_{xx} - C_z^{-1} C V_{xx} \right\}, & \text{if } \delta < \alpha < 1 \end{cases} \\
&= \begin{cases} f_{u_{0\tau}}(0) \left[-C_z^{-1} \left\{ \Omega_{xx} + \int dB_x J_x^{c\prime} + C \int J_x^c J_x^{c\prime} \right\} \right], & \text{if } \alpha = 1 \\ f_{u_{0\tau}}(0) \left[-C_z^{-1} \left\{ \Omega_{xx} + C V_{xx} \right\} \right], & \text{if } \delta < \alpha < 1 \end{cases} := f_{u_{0\tau}}(0)\, \Psi_{cxz}.
\end{aligned}
$$

and using the fact $dJ_x^c = dB_x + C J_x^c$ from Ito calculus:

$$
\int dB_x J_x^{c\prime} + C \int J_x^c J_x^{c\prime} = \int dJ_x^c J_x^{c\prime}.
$$

(ii) For (MI) and (I1) cases with $\alpha \in (0, \delta)$,

$$
\begin{aligned}
M_{\beta_\tau,n} &= \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, \tilde{Z}_{t-1,n} X'_{t-1,n} \\
&= \frac{1}{n^{1+\alpha}} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, \tilde{z}_{t-1} x'_{t-1} = f_{u_{0\tau}}(0) \left\{ \frac{1}{n^{1+\alpha}} \sum_{t=1}^{n} \tilde{z}_{t-1} x'_{t-1} \right\} + o_p(1) \text{ as earlier,} \\
&= f_{u_{0\tau}}(0) \left\{ \frac{1}{n^{1+\alpha}} \sum_{t=1}^{n} x_{t-1} x'_{t-1} + o_p(1) \right\} + o_p(1) \text{ (from MP}_b\text{, lemma 3.5-(ii))} \\
&\Longrightarrow f_{u_{0\tau}}(0)\, V_{xx}.
\end{aligned}
$$

(iii) If x_{t-1} belongs to (I0), using Lemma A.2 in Kostakis et al. (2012) again,

$$
M_{\gamma_\tau,n} = \frac{1}{n} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, \tilde{z}_{t-1} x'_{t-1} = \frac{1}{n} \sum_{t=1}^{n} f_{u_{0t\tau},t-1}(0)\, x_{t-1} x'_{t-1} + o_p(1) \to^p f_{u_{0\tau}}(0)\, \Omega_{xx}
$$

■

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

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

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 $n^{-\frac{1+\alpha}{2}}$ hence omitted. (I0) case is again standard, so omitted.

Proof for (ME) predictors will be discussed below (Lemma 7.4).

The stronger normalizer $n^{-\frac{1+\delta}{2}}$ (than $n^{-\frac{1}{2}}$) 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 the paper will hold here. For completeness, I provide the modified proof.

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

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

Proof. WLOG, we assume $K = 1$. I use the following standard argument:

$$\begin{aligned}
&\sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C\right\}} |G_n(\epsilon) - G_n(0)| \\
= &\sup_{\bigcup_{j=1,\dots,J}\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} |G_n(\epsilon) - G_n(0)| \\
= &\max_{j=1,\dots,J}\left(\sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} |G_n(\epsilon) - G_n(0)|\right) \\
\leq &\max_{j=1,\dots,J}\left(\sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} |G_n(\epsilon) - G_n(l_j)| + |G_n(l_j) - G_n(0)|\right) \\
= &\max_{j=1,\dots,J}\left(\sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} |G_n(\epsilon) - G_n(l_j)|\right) + \max_{j=1,\dots,J} |G_n(l_j) - G_n(0)|
\end{aligned}$$

where $\left\{|\epsilon| \leq n^{-(1+\delta)/2}C(\epsilon_j) : j = 1, ..., J\right\}$ is a collection of closed intervals with a length ϵ_j whose union is the original shrinking neighborhood $\left\{|\epsilon| \leq n^{-(1+\delta)/2}C\right\}$ (see below). l_j is the lowest vertex of each interval. Therefore, we show (i) $\max_{j=1,...,J} |G_n(l_j) - G_n(0)| = o_p(1)$, and (ii) $\max_{j=1,...,J}\left(\sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} |G_n(\epsilon) - G_n(l_j)|\right) = o_p(1)$.

(i) $\max_{j=1,...,J} |G_n(l_j) - G_n(0)| = o_p(1)$.

It suffices to show that

$$G_n(c_n) - G_n(0) = o_p(1) \tag{7.2}$$

for $c_n = \frac{c}{n^{\frac{1+\delta}{2}}}$ with any fixed $|c| \leq C$, hence a pointwise convergence. I will use the following argument: if $(\xi_{nt}, \mathcal{F}_t)$ is martingale difference array (mda), and the (predictable) quadratic variation $\sum_{t=1}^n E\left[\xi_{nt}\xi_{nt}' | \mathcal{F}_{t-1}\right] = o_p(1)$, then

$$\sum_{t=1}^n \xi_{nt} = o_p(1).$$

Let $\xi_{nt} := \frac{1}{n^{\frac{1+\delta}{2}}}\tilde{z}_{t-1}\zeta_t$ where

$$\zeta_t := \psi_\tau\left(u_{0t\tau} - c_n x_{t-1}\right) - E_{t-1}\left[\psi_\tau\left(u_{0t\tau} - c_n x_{t-1}\right)\right] - \psi_\tau\left(u_{0t\tau}\right) + E_{t-1}\left[\psi_\tau\left(u_{0t\tau}\right)\right]$$

then

$$G_n\left(c_n\right) - G_n\left(0\right) = \sum_{t=1}^{n} \xi_{nt}$$

and by construction ζ_t are mds and ξ_{nt} are mda.

Note that

$$\zeta_t = 1\left(u_{0t\tau} < 0\right) - 1\left(u_{0t\tau} < c_n x_{t-1}\right) + \Pr\left(u_{0t\tau} < c_n x_{t-1}|\mathcal{F}_{t-1}\right) - \Pr\left(u_{0t\tau} < 0|\mathcal{F}_{t-1}\right).$$

Given $\mathcal{F}_{t-1}$, (a) assume $c_n x_{t-1} > 0$, then

$$\zeta_t = -1\left(0 \le u_{0t\tau} < c_n x_{t-1}\right) + \Pr\left(0 \le u_{0t\tau} < c_n x_{t-1}|\mathcal{F}_{t-1}\right),$$

and

$$\begin{aligned}
\zeta_t^2 &= 1\left(0 \le u_{0t\tau} < c_n x_{t-1}\right) + \left\{\Pr\left(0 \le u_{0t\tau} < c_n x_{t-1}|\mathcal{F}_{t-1}\right)\right\}^2 \\
&\quad -2\Pr\left(0 \le u_{0t\tau} < c_n x_{t-1}|\mathcal{F}_{t-1}\right) 1\left(0 \le u_{0t\tau} < c_n x_{t-1}\right),
\end{aligned}$$

hence

$$\begin{aligned}
E\left[\zeta_t^2|\mathcal{F}_{t-1}\right] &= \left\{\Pr\left(0 \le u_{0t\tau} < c_n x_{t-1}|\mathcal{F}_{t-1}\right)\right\}\left\{1 - \Pr\left(0 \le u_{0t\tau} < c_n x_{t-1}|\mathcal{F}_{t-1}\right)\right\} \\
&\le \left\{\Pr\left(0 \le u_{0t\tau} < c_n x_{t-1}|\mathcal{F}_{t-1}\right)\right\} \le f_{u_{0t\tau},t-1}(0)c_n x_{t-1} + o_p(1) \\
&= O_p\left(c_n\sqrt{n}\right) = O_p\left(\frac{1}{n^{\frac{\delta}{2}}}\right) \text{ for all } t,
\end{aligned}$$

thus $\sup_t E\left[\zeta_t^2|\mathcal{F}_{t-1}\right] = o_p(1)$.

It is easy to show for the case (b) $c_n x_{t-1} \le 0$ given $\mathcal{F}_{t-1}$ in a similar way. Therefore

$$\begin{aligned}
\sum_{t=1}^{n} E\left[\xi_{nt}^2|\mathcal{F}_{t-1}\right] &= \frac{1}{n^{1+\delta}}\sum_{t=1}^{n}\tilde{z}_{t-1}^2 E\left[\zeta_t^2|\mathcal{F}_{t-1}\right] \\
&\le_{a.s} \left(\sup_t E\left[\zeta_t^2|\mathcal{F}_{t-1}\right]\right)\left(\frac{1}{n^{1+\delta}}\sum_{t=1}^{n}\tilde{z}_{t-1}^2\right) = o_p(1),
\end{aligned}$$

confirming (7.2).

(ii) $\max_{j=1,\ldots,J}\left(\sup_{\left\{|\epsilon| \le n^{-(1+\delta)/2}C(\epsilon_j)\right\}}\left|G_n\left(\epsilon\right) - G_n\left(l_j\right)\right|\right) = o_p(1).$

For each j,

$$
\begin{aligned}
&\sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} \left|G_n\left(\epsilon\right)-G_n\left(l_j\right)\right| \\
=& \sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} \left|\frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n}\tilde{z}_{t-1}\left\{1\left(l_jx_{t-1}\leq u_{0t\tau}<\epsilon x_{t-1}\right)-\Pr\left(l_jx_{t-1}\leq u_{0t\tau}<\epsilon x_{t-1}|\mathcal{F}_{t-1}\right)\right\}\right| \\
\leq& \sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} \frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n}\left|\tilde{z}_{t-1}\right|1\left(l_jx_{t-1}\leq u_{0t\tau}<\epsilon x_{t-1}\right) \\
&+ \sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} \frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n}\left|\tilde{z}_{t-1}\right|\Pr\left(l_jx_{t-1}\leq u_{0t\tau}<\epsilon x_{t-1}|\mathcal{F}_{t-1}\right).
\end{aligned}
$$

Now since $1\left(l_jx_{t-1}\leq u_{0t\tau}<\epsilon x_{t-1}\right)\leq 1\left(-n^{-(1+\delta)/2}\epsilon_jx_{t-1}\leq u_{0t\tau}<n^{-(1+\delta)/2}\epsilon_jx_{t-1}\right)$ from the monotonicity,

$$
\begin{aligned}
&\sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} \left(\frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n}\left|\tilde{z}_{t-1}\right|1\left(l_jx_{t-1}\leq u_{0t\tau}<\epsilon x_{t-1}\right)\right) \\
\leq& \frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n}\left|\tilde{z}_{t-1}\right|1\left(-n^{-(1+\delta)/2}\epsilon_jx_{t-1}\leq u_{0t\tau}<n^{-(1+\delta)/2}\epsilon_jx_{t-1}\right).
\end{aligned}
$$

Let $\xi_{nt}=\frac{1}{n^{\frac{1+\delta}{2}}}\left|\tilde{z}_{t-1}\right|1\left(-n^{-(1+\delta)/2}\epsilon_jx_{t-1}\leq u_{0t\tau}<n^{-(1+\delta)/2}\epsilon_jx_{t-1}\right)$, then

$$
\begin{aligned}
&\sum_{t=1}^{n}E_{t-1}\left[\xi_{nt}\right]=\frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n}\left|\tilde{z}_{t-1}\right|\Pr\left(-n^{-(1+\delta)/2}\epsilon_jx_{t-1}\leq u_{0t\tau}<\epsilon_jn^{-(1+\delta)/2}x_{t-1}|\mathcal{F}_{t-1}\right) \\
\leq& \frac{2M_f}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n}\left|\tilde{z}_{t-1}\right|\left|\epsilon_jn^{-(1+\delta)/2}x_{t-1}\right|=O_p\left(\epsilon_j\right),
\end{aligned}
$$

uniformly in ϵ_j, where we use that $f_{u_{0t\tau},t-1}(\cdot)$ is bounded above a.s (in the local neighborhood of zero) by M_f from Assumption 2.1-(ii).

We can also show

$$
\sum_{t=1}^{n}E_{t-1}\left[\xi_{nt}^2\right]=o_p(1),
$$

as before. Thus,

$$
\sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} \left(\frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n}\left|\tilde{z}_{t-1}\right|1\left(l_jx_{t-1}\leq u_{0t\tau}<\epsilon x_{t-1}\right)\right)=o_p(1),
$$

Similarly we can show

$$
\sup_{\left\{|\epsilon|\leq n^{-(1+\delta)/2}C(\epsilon_j)\right\}} \left(\frac{1}{n^{\frac{1+\delta}{2}}}\sum_{t=1}^{n}\left|\tilde{z}_{t-1}\right|\Pr\left(l_jx_{t-1}\leq u_{0t\tau}<\epsilon x_{t-1}|\mathcal{F}_{t-1}\right)\right)=o_p(1),
$$

confirming (ii). ■

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 $n^{-\frac{1+\delta}{2}}$. Hence, we can show $\left(\hat{\beta}_{1,\tau}^{IVXQR} - \beta_{1,\tau}\right) = O_p\left(n^{-\frac{1+\delta}{2}}\right)$. Let $\hat{\beta}_{1,\tau} = \hat{\beta}_{1,\tau}^{IVXQR}$ within this proof.

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

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

With notation of embedded normalizers,

$$o_p(1) = \sum_{t=1}^{n} \left\{ \tilde{Z}_{t-1,n} \psi_\tau\left(u_{0t\tau}\right) + \tilde{Z}_{t-1,n} E_{t-1}\left(\psi_\tau\left(u_{0t\tau} - \hat{\epsilon}_\tau' x_{t-1}\right)\right)\right\}, \tag{7.3}$$

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 (7.3),

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

therefore,

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

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),*

$$\frac{1}{n^{(\alpha\wedge\delta)}}R_n^{-n}\sum_{t=1}^{n}\tilde{z}_{t-1}\psi_\tau\left(u_{0t\tau}\right)\Longrightarrow CC_{z\alpha\delta}\times N\left(0,\tau(1-\tau)\tilde{V}_{xx}\right),$$

and

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

where

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

Proof. The result directly follows from the proof of Lemma 5.4 and 5.5 in PL$_b$, by replacing u_{0t} with $\psi_\tau\left(u_{0t\tau}\right)$. Thus, the IVX-QR limit theory in Theorem 3.1 for (ME) case follows by the same procedure to the earlier proofs - (MI) and (I1) cases. ■

Proof of Theorem 3.2. Note that

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

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

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

where

$$G_{\tau,n}^*=\sum_{t=1}^{n}\tilde{Z}_{t-1,n}\psi_\tau\left(u_{0t\tau}^*\right),$$

and it is clear that $G_{\tau,n}^*=G_{\tau,n}$ under $H_0:\beta_{1,\tau}=0$, leading to

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

■

7.5 Local Power with Estimated Intercepts

Figure 5: $c = -5$ ($n = 250$, $R = 0.98$) with t(4), t(3), t(2) and t(1) innovations.

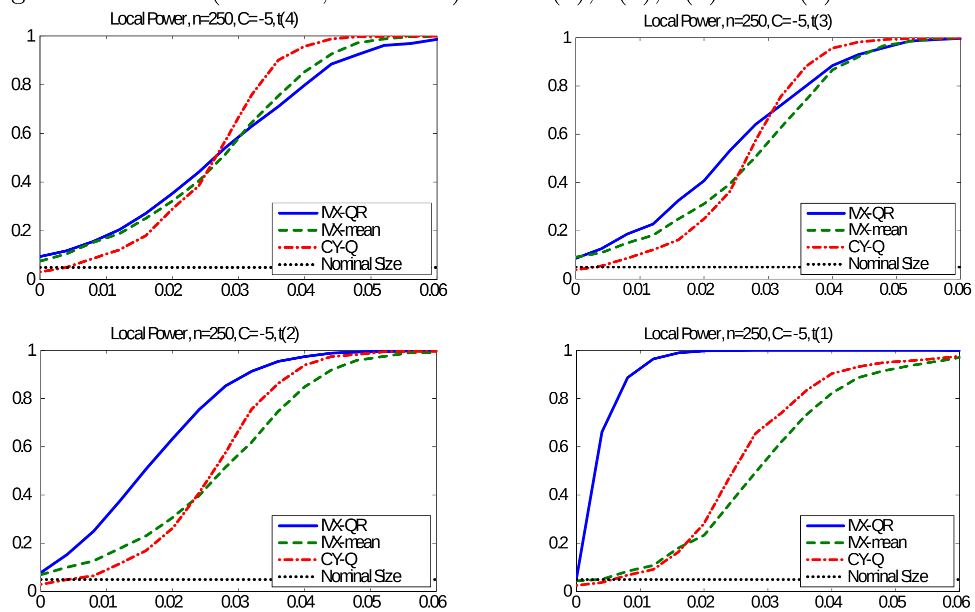
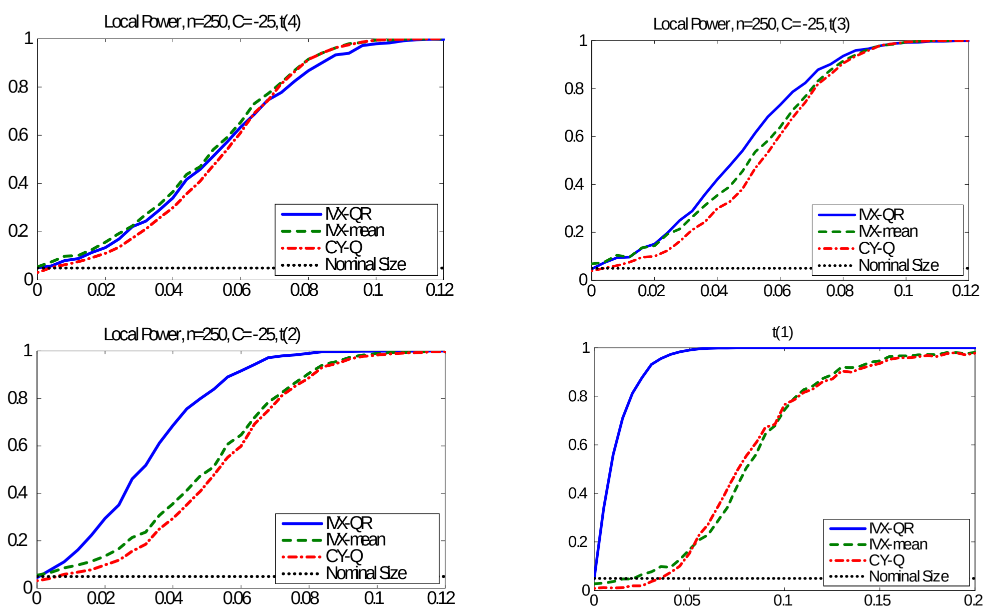


Figure 6: $c = -25$ ($n = 250$, $R = 0.9$) with t(4), t(3), t(2) and t(1) innovations.



7.6 Diagnostic Tests for Comovement between Persistent Predictors

Table 9: Diagnostic Tests for Comovement (full sample period: Jan 1927 to Dec 2005)

(d/p, tbl)	correlation	Johansen tests for cointegration		
	-0.16	maximum rank	trace statistics	5% critical value
		0	11.76*	15.41
		1	3.46	3.76
(d/p, dfy)	correlation	Johansen tests for cointegration		
	0.44	maximum rank	trace statistics	5% critical value
		0	20.74	15.41
		1	2.99*	3.76
(d/e, tbl)	correlation	Johansen tests for cointegration		
	-0.52	maximum rank	trace statistics	5% critical value
		0	21.54	15.41
		1	7.46	3.76
(d/e, dfy)	correlation	Johansen tests for cointegration		
	0.56	maximum rank	trace statistics	5% critical value
		0	33.98	15.41
		1	7.57	3.76
(d/p, b/m)	correlation	Johansen tests for cointegration		
	0.82	maximum rank	trace statistics	5% critical value
		0	23.74	15.41
		1	3.82	3.76
(d/e, b/m)	correlation	Johansen tests for cointegration		
	0.29	maximum rank	trace statistics	5% critical value
		0	21.82	15.41
		1	6.78	3.76
(tbl, b/m)	correlation	Johansen tests for cointegration		
	0.07	maximum rank	trace statistics	5% critical value
		0	18.37	15.41
		1	7.20	3.76
(dfy, b/m)	correlation	Johansen tests for cointegration		
	0.52	maximum rank	trace statistics	5% critical value
		0	25.79	15.41
		1	8.64	3.76

Table 10: Diagnostic Tests for Comovement (subperiod: Jan 1952 to Dec 2005)

(b/m, d/p)	correlation	Johansen tests for cointegration		
	0.88	maximum rank	trace statistics	5% critical value
		0	10.34*	15.41
		1	2.85	3.76
(b/m, dfy)	correlation	Johansen tests for cointegration		
	0.47	maximum rank	trace statistics	5% critical value
		0	32.86	15.41
		1	3.13*	3.76
(b/m, e/p)	correlation	Johansen tests for cointegration		
	0.89	maximum rank	trace statistics	5% critical value
		0	9.70*	15.41
		1	1.95	3.76
(b/m, tbl)	correlation	Johansen tests for cointegration		
	0.50	maximum rank	trace statistics	5% critical value
		0	14.89*	15.41
		1	2.00	3.76
(d/p, tbl)	correlation	Johansen tests for cointegration		
	0.35	maximum rank	trace statistics	5% critical value
		0	15.68	15.41
		1	2.05*	3.76
(dfy, tbl)	correlation	Johansen tests for cointegration		
	0.58	maximum rank	trace statistics	5% critical value
		0	40.87	15.41
		1	5.77	3.76
(e/p, tbl)	correlation	Johansen tests for cointegration		
	0.53	maximum rank	trace statistics	5% critical value
		0	16.74	15.41
		1	3.75*	3.76

8 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.

Cont, R (2001), "Empirical properties of asset returns: stylized facts and statistical issues", *Quantitative Finance*, 1:2, 223-336

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.

Embrechts, P., Kluppelberg, C. and Mikosch, T (1997). *Modelling Extremal Events*, Springer Verlag, Berlin.

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

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.

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

Jansson, M. and M. Moreira (2006). "Optimal inference in regression models with nearly integrated regressors," *Econometrica*, 74(3), 681-714.

Johansen, S (1988), "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12(2-3), pp. 231-254.

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 (2010). "Robust econometric inference for stock return predictability," *Unpublished Manuscript,* University of Nottingham.

Magdalinos, T. and P. C. B Phillips (2009a). “Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors”. *Econometric Theory*, 25, 482-526.

Magdalinos, T. and P. C. B Phillips (2009b). “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.*

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 (2012), "On confidence Intervals for autoregressive roots and predictive regression," *Unpublished Manuscript*.

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 (2012a), "Predictive regression under various degrees of persistence and robust long-horizon regression," *Unpublished Manuscript*.

Phillips, P. C. B and J.H. Lee (2012b), "Robust econometric inference with mixed integrated and mildly explosive regressors," *Unpublished Manuscript*.

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

Phillips, P. C. B. and T. Magdalinos (2007b), "Limit theory for moderate deviations from a unit root under weak dependence," in G. D. A. Phillips and E. Tzavalis (Eds.) *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis.* Cambridge: Cambridge University Press, pp.123-162.

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.