Exact Inference in Long-Horizon Predictive Quantile Regressions with an Application to Stock Returns*

Sermin Gungor[1] **and Richard Luger**[2]

[1]University of Western Ontario, [1]Bank of Canada, and [2]Laval University

Address correspondence to Richard Luger, Department of Finance, Insurance and Real Estate, Laval University, Quebec City, Quebec G1V 0A6, Canada, or e-mail: richard.luger@fsa.ulaval.ca.

Abstract

We develop an exact and distribution-free procedure to test for quantile predictability at several prediction horizons and quantile levels *jointly*, while allowing for an endogenous predictive regressor with any degree of persistence. The approach proceeds by combining together the quantile regression t-statistics from each considered prediction horizon and quantile level, and uses Monte Carlo resampling techniques to control the familywise error rate in finite samples. A simulation study confirms that the proposed inference procedure is indeed level-correct and that testing several quantile levels jointly can deliver more power to detect predictability. In an empirical application to excess stock returns, we find that the default yield spread predicts the right tail while the short-term interest rate predicts the centre of the return distribution. This predictability evidence is stronger at shorter rather than longer horizons.

Key words: predictability, quantile regression, multiple comparisons, persistent predictor, Monte Carlo permutation test, exact distribution-free inference

JEL classification: C14, C22, G17

*We would like to thank two anonymous reviewers, an associate editor, and the editor (Federico Bandi) for their comments and suggestions. We are also thankful for useful discussions with conference and seminar participants at the 2016 International Conference on Computational and Financial Econometrics, the 2017 International Conference on Computing in Economics and Finance, the 2017 Annual Meeting of the Southern Finance Association, the 3rd Conference on Financial Econometrics and Risk Management, le 57e congrès annuel de la Société canadienne de science économique, the 2017 Meeting of the Canadian Econometric Study Group (Poster Session), la Journée du CRREP 2018, and CREST. Richard Luger gratefully acknowledges financial support by the Social Sciences and Humanities Research Council of Canada. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Canada.

1 Introduction

In a traditional predictive regression setting, an outcome variable of interest is regressed on the lagged value of a predictor variable. The relationship under investigation is

$$E(y_t \mid x_{t-1}) = \beta_0 + \beta_1 x_{t-1}, \tag{1}$$

where y_t is the outcome variable and x_t is the predictor variable. If $\beta_1 \neq 0$, then x_t is said to have predictive ability for the mean of the outcome distribution. The econometric method used to detect such predictability is ordinary least squares (OLS) and the null hypothesis of no predictability, *i.e.,* $\beta_1 = 0$ in (1), is tested with a t-statistic. It is well known since the work by Mankiw and Shapiro (1986) and Stambaugh (1999) that the usual asymptotic theory provides a poor approximation to the finite-sample distribution of the t-statistic, resulting in far too many spurious rejections of the null hypothesis. This problem of false positives worsens as the predictor variable x_t becomes more persistent and its errors are more correlated with those of the regression model. A leading example is stock return predictability using a highly persistent regressor, like the dividend-price ratio, the earnings-price ratio, the book-to-market ratio, and various interest rates and interest rate spreads.

A common belief in the literature is that such predictive regressions of stock returns typically contain more noise than signal coming from x_{t-1}, and that the signal-to-noise ratio can be strengthened by aggregating the outcome variable forwards. For a given prediction horizon $h \geq 1$, the aggregated outcome variable is constructed as

$$y_t^h = \sum_{k=0}^{h-1} y_{t+k},$$

which is then regressed onto x_{t-1}. By construction, the time series of overlapping observations $\{y_t^h \,|\, h > 1\}$ contains a moving-average error structure. Examples of these types of long-horizon predictive regressions include Campbell and Shiller (1988), Fama and French (1988), Lanne (2002), Valkanov (2003), and Hjalmarsson (2011).[1]

A recent strand of the literature uses the quantile regression method of Koenker and Bassett (1978) to examine whether the outcome distribution is predictable not just in the centre, but also in the shoulders and tails; see Cenesizoglu and Timmermann (2008), Maynard et al. (2011) and Lee (2016). In addition to offering a more complete view of the outcome distribution, quantile-based methods are attractive for many other reasons. First of all, quantiles are always defined whereas moments may not exist (*e.g.*, the moments of the Cauchy distribution are all undefined). It is also well known that the sample mean (and its regression version, the OLS estimator) is very sensitive to outliers. And conventional measures of higher moments (*e.g.*, skewness and kurtosis) are also sensitive to the presence of outliers, since they are based on sample averages. The median and interquartile range are well known as robust measures of location and dispersion. Quantiles can even be used to construct robust alternative measures of skewness and kurtosis; see Kim and White (2004) for instance. Bassett et al. (2004) show that a general form of pessimistic portfolio optimization may be formulated as a problem of linear quantile regression. In so doing they provide an interesting link between quantile regression and expected shortfall (ES), a risk measure that builds on the standard (quantile-based) value-at-risk (VaR) measure of market risk. Taylor (2017) further exploits that link to model VaR and ES, jointly.

So instead of the conditional expectation, the studies by Cenesizoglu and Timmermann

[1] There is also work that attempts to boost the signal by aggregating the predictor variable x_t either backwards or forwards; see, for example, Boudoukh and Richardson (1993), Valkanov (2003), Bandi and Perron (2008), and Bandi et al. (2018).

(2008), Maynard et al. (2011) and Lee (2016) focus on the quantile regression

$$Q_\tau(y_t \,|\, x_{t-1}) = \beta_0(\tau) + \beta_1(\tau)x_{t-1}, \tag{2}$$

where $Q_\tau(y_t \,|\, x_{t-1})$ is the conditional quantile of y_t at a given quantile level $\tau \in (0, 1)$. An emerging view is that many economic variables seem to have far greater predictive ability for the outer quantiles rather than for the centre of the stock return distribution. A potential pitfall, however, is that the standard quantile regression t-statistic is also prone to the overrejection problem, just like its mean regression counterpart; see Lee (2016) and the simulation results presented here below in Section 4.

Maynard et al. (2011) address the size distortion problem by deriving the asymptotic distribution of the quantile regression t-statistic under a local-to-unity specification for the predictor variable. Lee (2016) adopts the so-called extended instrumental variable (IVX) methodology of Magdalinos and Phillips (2009) to deal with the presence of persistent regressors in predictive quantile regressions. The idea underlying the IVX methodology is to generate an instrumental variable of intermediate persistence by filtering a persistent and possibly endogenous regressor; Kostakis et al. (2015) extend this approach to a multiple predictor context. Lee (2016) suggests using the instrument in lieu of the original predictor in the quantile regression and derives the asymptotic distribution of the resulting estimator under the null of no quantile predictability, *i.e.*, $\beta_1(\tau) = 0$ in (2). He calls this approach IVX-QR.

What is common to the large-sample approaches in Maynard et al. (2011) and Lee (2016) is that they only test for quantile predictability at a single quantile level, τ. Under the null hypothesis of no predictability, however, the outcome distribution should not be predictable at *any* quantile level. Yet by testing for predictability at enough quantile levels, false positives will occur. The multiple comparisons problem then consists of combining the predictability test results from each considered quantile level, say $\tau_1, ..., \tau_q$, in such a way that controls the familywise error rate in finite samples; see Hsu (1996) for the theory of multiple comparisons. In a long-horizon setting, the multiple comparisons problem may also include several prediction horizons $h_1, ..., h_p$. A familywise error rate (FWER) can be defined as

$$\text{FWER} = \Pr\big(\text{Reject at least one } H_0(h_i, \tau_j) \,|\, \text{all } H_0(h_i, \tau_j) \text{ are true}\big),$$

where $H_0(h_i, \tau_j)$ refers to the null hypothesis of no predictability at the specific prediction horizon h_i and quantile level τ_j, *i.e.*, $\beta_1(h_i, \tau_j) = 0$ in

$$Q_{\tau_j}(y_t^{h_i} \,|\, x_{t-1}) = \beta_0(h_i, \tau_j) + \beta_1(h_i, \tau_j) x_{t-1}.$$

In words, the FWER is the probability of making at least one Type I error. Given a desired significance level α, the challenge taken up here is to devise a procedure to test the absence of quantile predictability such that FWER $\leq \alpha$, no matter the sample size.

Consider for a moment the one-period-ahead case. Given a finite-sample test for single-quantile predictability at horizon $h = 1$, one could then of course control FWER($h = 1$) = Pr(Reject at least one $H_0(\tau_j) \,|\,$ all $H_0(\tau_j)$ are true) at level α by applying the familiar Bonferroni procedure which consists of rejecting any hypothesis $H_0(h = 1, \tau_i) = H_0(\tau_i)$ with p-value $p(\tau_i) \leq \alpha/q$, where q is the number of quantiles being tested. The corresponding Bonferroni-adjusted p-value is given by $\min(q \times p(\tau_i), 1)$. Cenesizoglu and Timmermann

(2008) make the Bonferroni adjustment to (asymptotically justified) bootstrapped p-values in order to obtain a joint test across all considered quantiles which is robust to arbitrary dependencies among the individual p-values. The problem with this approach is that it will be lacking in power as q grows. Instead of the Bonferroni adjustment, Westfall and Young (1993) advocate resampling-based methods to obtain less conservative multiple testing procedures which take into account the dependence structure between test statistics. The idea that resampling can be used to estimate the joint distribution of p-values is central to resampling-based testing.

In this spirit, we propose an *exact* resampling-based procedure for controlling the joint significance of quantile predictability tests at several quantile levels and prediction horizons. We achieve control of the FWER by exploiting a permutation principle which holds under the null hypothesis of no quantile predictability, assuming that the predictor variable evolves according to an AR(1) model – a first-order autoregressive model. The AR(1) model is a standard assumption in the mean predictability literature; see Cavanagh et al. (1995), Stambaugh (1999), Lewellen (2004), Torous et al. (2004), Amihud and Hurvich (2004), Campbell and Yogo (2006), and Polk et al. (2006) among others. Our approach is attractive for a number of reasons: (i) it places no restrictions on the persistence of the predictor variable, thereby allowing for unit-root and even explosive behaviour; (ii) it is invariant to the contemporaneous dependence between the errors of the predictor model and those of the predictive regression model; (iii) it makes no parametric distributional assumptions, thereby allowing for heavy-tailed errors; and (iv) it accounts for the overlapping observations used in long-horizon regressions.

The developed approach rests on an equally likely property that holds for the collection of quantile regression t-statistics together with the AR(1) regression t-statistic. This key property paves the way for an exact distribution-free test procedure using the technique of Monte Carlo (MC) tests (Dwass, 1957; Barnard, 1963; Birnbaum, 1974). Specifically, the proposed test procedure proceeds by linking: (i) each point in an exact confidence set for the AR(1) parameter, with (ii) MC p-values of a test statistic for quantile predictability. The confidence set is obtained by collecting all the values that cannot be rejected by an MC t-ratio test for the AR(1) parameter, while the predictability test statistic combines the quantile regression t-statistics at each considered prediction horizon and quantile level of the outcome distribution. This simultaneous inference approach yields a test of no quantile predictability that controls the FWER in finite samples.

As an illustration, we apply the developed test procedure to the distribution of excess returns on the S&P value-weighted stock market index. We uncover evidence of quantile predictability using interest rate variables, while valuation ratios seem to have little to no predictive ability. Specifically, the default yield predicts the right-tail quantiles and the short-term interest rate predicts the central quantiles of the return distribution. This evidence appears strongest at short horizons up to three months and disappears as the prediction horizon grows longer.

The rest of the paper is organized as follows. Section 2 describes the statistical framework and Section 3 develops the long-horizon quantile predictability test procedure. Section 4 reports the results of a series of simulation experiments designed to examine the empirical size and power of the test procedure, as well as its robustness to departures from the maintained assumptions. The results confirm the fact that the proposed inference procedure is level-correct and that testing several quantile levels jointly (instead of individually) can yield more power to detect predictability, especially when the predictor is persistent over time. Section 5 presents the empirical application and a discussion of the findings. Section 6 concludes.

2 Statistical Framework

The setting involves the outcome variable of interest y_t and another variable x_{t-1}, observed at time $t-1$, which could have the ability to predict y_t. More precisely, we work within a framework involving the collection of random variables $y_1, ..., y_T$, $x_0, x_1, ..., x_T$ and we say that x_{t-1} does not predict y_t if the following condition holds:

$$y_t \text{ is independent of } x_{t-1}, \text{ for each } t = 1, ..., T. \tag{3}$$

Let $F(y_t)$ and $F(y_t \,|\, x_{t-1})$ denote, respectively, the unconditional distribution and the conditional distribution of y_t, given x_{t-1}. Condition (3) can then be stated as

$$F(y_t \,|\, x_{t-1}) = F(y_t), \text{ for each } t = 1, ..., T,$$

while, under the alternative hypothesis of predictability, x_{t-1} affects some part of the outcome distribution. In order to complete the statistical framework, a time-series model for x_t will be specified. As typically done in the mean predictability literature, x_t will be assumed to evolve according to an AR(1) model.

Let $\boldsymbol{\eta}_t = (u_t, v_t)'$ be a vector of random error terms. We formally state the null hypothesis of no predictability as

$$H_0 : \begin{cases} y_t &= \beta_0 + u_t, \\ x_t &= \mu + \phi x_{t-1} + v_t, \end{cases} \tag{4}$$

where $\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T$ is a sequence of exchangeable random vectors, and where x_0 is either fixed or random but independent of $\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T$. The exchangeability requirement means that, for every permutation $d_1, ..., d_T$ of the integers $1, ..., T$, we have

$$(\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T) \stackrel{d}{=} (\boldsymbol{\eta}_{d_1}, ..., \boldsymbol{\eta}_{d_T}), \tag{5}$$

where the symbol '$\stackrel{d}{=}$' stands for the equality in distribution (Randles and Wolfe, 1979, Definition 1.3.6). Note that the elements within $\boldsymbol{\eta}_t$ need not be exchangeable. If the vectors $\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T$ are independent and identically distributed (i.i.d.), then they are obviously exchangeable.[2] On the other hand, exchangeable random vectors are not necessarily independent. For example, stack the error vectors into the $T \times 2$ matrix $\boldsymbol{\eta} = [\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T]'$ and suppose $\boldsymbol{\eta}$ follows the matrix normal distribution, denoted by $\boldsymbol{\eta} \sim \mathcal{N}_{T,2}(\boldsymbol{M}, \boldsymbol{\Xi}, \boldsymbol{\Sigma})$. In this $T \times 2$ case, the associated density function is given by

$$f(\boldsymbol{\eta}) = \frac{1}{(2\pi)^T |\boldsymbol{\Xi}||\boldsymbol{\Sigma}|^{T/2}} \exp\left\{ -\frac{1}{2} tr[\boldsymbol{\Xi}^{-1}(\boldsymbol{\eta} - \boldsymbol{M})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta} - \boldsymbol{M})'] \right\},$$

where $\boldsymbol{M}$, $\boldsymbol{\Xi}$, and $\boldsymbol{\Sigma}$ are $T \times 2$, $T \times T$, and 2×2 matrices, respectively. Clearly, $\boldsymbol{M}$ is the matrix-valued expected value of $\boldsymbol{\eta}$. The left matrix $\boldsymbol{\Xi}$ represents the covariance between the rows of $\boldsymbol{\eta}$ and the right matrix $\boldsymbol{\Sigma}$ represents the covariance between the columns of $\boldsymbol{\eta}$. If $\boldsymbol{M} = \boldsymbol{0}$ and $\boldsymbol{\Xi} = (1 - \rho)\boldsymbol{I}_T + \rho \boldsymbol{J}_T$, where ρ is an equicorrelation parameter, $\boldsymbol{I}_T$ denotes

[2]The i.i.d. assumption is often maintained in the stock return predictability literature; see, for instance, Stambaugh (1999), Amihud and Hurvich (2004), and Campbell and Yogo (2006), among others.

the T-dimensional identity matrix and $\boldsymbol{J}_T$ is a $T \times T$ matrix of ones, then $f([\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T]') = f([\boldsymbol{\eta}_{d_1}, ..., \boldsymbol{\eta}_{d_T}]')$ for any permutation $d_1, ..., d_T$ of $1, ..., T$. That is, $\boldsymbol{\eta}_1, ..., \boldsymbol{\eta}_T$ are exchangeable, yet they are not independent. See Gupta and Nagar (2000) for a thorough treatment of the matrix normal distribution.

Besides exchangeability, no other distributional assumption is made, thereby leaving open the possibility of heavy-tailed $\boldsymbol{\eta}_t$ errors. The parameters β_0, μ, and ϕ appearing in (4) are unknown, and $\phi \in \mathcal{D}_\phi$, where $\mathcal{D}_\phi \subseteq \mathbb{R}$ is a nonempty set of admissible values for ϕ. Depending on the context, the choice of $\mathcal{D}_\phi$ could restrict ϕ to stationary values, or it could allow for a unit root and even explosive values. Indeed, the set $\mathcal{D}_\phi$ may be $\mathbb{R}$ itself, the open interval $(-1, 1)$, the closed interval $[-1, 1]$, or any other appropriate subset of $\mathbb{R}$.

Following Cenesizoglu and Timmermann (2008), Maynard et al. (2011), and Lee (2016), we use linear quantile regression models to detect predictability at various points of the outcome distribution. With the long-horizon variable y_t^h, these models are written here again as

$$Q_\tau(y_t^h \,|\, x_{t-1}) = \beta_0(h, \tau) + \beta_1(h, \tau) x_{t-1}, \tag{6}$$

where $Q_\tau(y_t^h \,|\, x_{t-1})$ is the conditional quantile of y_t^h at a given quantile level $\tau \in (0, 1)$. Stack the regressors appearing in (6) in the vector $\boldsymbol{x}_{t-1} = (1, x_{t-1})'$ and let $\boldsymbol{\theta}(h, \tau) = \big(\beta_0(h, \tau), \beta_1(h, \tau)\big)'$ be the vector comprising the associated horizon- and quantile-specific intercept and slope coefficients. The standard quantile regression coefficient estimates are then given by

$$\hat{\boldsymbol{\theta}}(h, \tau) = \underset{\boldsymbol{\theta}(h,\tau)}{\arg\min} \sum_{t=1}^{T} \rho_\tau\big(y_t^h - \boldsymbol{\theta}(h, \tau)' \boldsymbol{x}_{t-1}\big),$$

where $\rho_\tau(u) = (\tau - \mathbb{I}[u < 0])u$ is the quantile regression loss function. Here $\mathbb{I}[A]$ is the indicator function which equals 1 when event A occurs, and 0 otherwise. In matrix form, the employed regressors are $\boldsymbol{X}_{-1} = [\boldsymbol{x}_0, ..., \boldsymbol{x}_{T-1}]'$.

Like in conventional quantile regression inference (Koenker and Bassett, 1978), we consider the t-statistic

$$t(h, \tau) = \frac{\hat{\beta}_1(h, \tau)}{\sqrt{\widehat{\text{var}}\big(\hat{\beta}_1(h, \tau)\big)}}. \tag{7}$$

Here $\hat{\beta}_1(h, \tau)$ is the second element of $\hat{\boldsymbol{\theta}}(h, \tau)$ and $\widehat{\text{var}}\big(\hat{\beta}_1(h, \tau)\big)$ is the $(2, 2)$-entry in the matrix $\omega^2(\tau)(\boldsymbol{X}'_{-1}\boldsymbol{X}_{-1})^{-1}$, where $\omega^2(\tau) = \tau(1 - \tau)/f_u^2(F_u^{-1}(\tau))$ and $f_u(F_u^{-1}(\tau))$ denotes the density of u_t evaluated at the τth quantile. This resembles the usual t-statistic based on OLS except that the definition of the variance uses $\omega^2(\tau)$ instead of σ^2, the variance of u_t. We compute (7) with the `rq` command available with the R 'quantreg' package.[3] Note that (7) as such does not acknowledge the moving-average errors in y_t^h. As we will see later, the error structure (the overlap) is accounted for by the proposed MC procedure.

Observe that (4) entails a joint hypothesis as it states that all points of y_t's distribution are unaffected by x_{t-1}. And this is also true at longer prediction horizons $h > 1$, since $y_t^h = h\beta_0 + \sum_{k=0}^{h-1} u_{t+k}$ under the null hypothesis. Let $h_1, ..., h_p$ and $\tau_1, ..., \tau_q$ be the prediction horizons and quantile levels that will be used to test for the presence of predictability. Furthermore, let ϕ_0 be a specified value such that $\phi_0 \in \mathcal{D}_\phi$ and consider the following subhypothesis:

$$H_0(\phi_0) : \bigcap_{i=1}^{p} \bigcap_{j=1}^{q} \beta_1(h_i, \tau_j) = 0, \; \phi = \phi_0, \tag{8}$$

which states zero slope coefficients in the quantile regressions at each considered prediction

[3]Specifically, we obtain the statistic in (7) with the `summary` function using the `iid` option. This presumes that the errors are i.i.d. and computes a scalar sparsity (*i.e.*, the reciprocal of the error density at the τth quantile) estimate using the Hall-Sheather bandwidth rule (Hall and Sheather, 1988) for the Siddiqui (1960) estimator (cf. Koenker, 2005, §3.4).

horizon and quantile level, and an admissible value for x_t's persistence parameter.

The joint null hypothesis of no predictability at prediction horizons $h_1, ..., h_p$ and quantile levels $\tau_1, ..., \tau_q$ can then be viewed as

$$H_0 : \bigcup_{\phi_0 \in \mathcal{D}_\phi} H_0(\phi_0), \tag{9}$$

where the union is taken over all admissible values for the AR(1) parameter in (4). Note that the number of prediction horizons and quantile levels that can be considered is restricted by the sample size. Which ones to consider depends on one's *a priori* beliefs about predictability. If one has no beliefs about which quantiles may be predictable, then a sensible approach is to perform a joint test over multiple quantile levels evenly spaced across the entire outcome distribution. Our simulation experiments in Section 4 show that such an approach can yield a more powerful predictability test, particularly as ϕ increases.

To close this section, we mention that Linton and Whang (2007) develop a method to test the absence of *own* predictability. This includes a test of $\beta_1(\tau) = 0$ in

$$Q_\tau(y_t \,|\, y_{t-1}) = \beta_0(\tau) + \beta_1(\tau) y_{t-1},$$

which is recognized as a first-order quantile autoregression (Koenker and Xiao, 2006). Our framework also yields a test of own predictability by setting $\phi = 0$ and $x_t = y_t$ in (4).

3 Exact Inference Methods

3.1 Test of $H_0(\phi_0)$

The t-statistic in (7) evaluated at each prediction horizon and quantile level appearing in (8) yields $t(h_1, \tau_1), ..., t(h_p, \tau_q)$. In order to test $H_0(\phi_0)$, these statistics can be combined via their maximum absolute value:

$$\mathcal{S} = \max_{1 \le i \le p} \max_{1 \le j \le q} |t(h_i, \tau_j)|,$$

which presumes two-sided alternatives. This approach corresponds to the *Studentized maximum modulus* method, described in Hsu (1996). It is also related to the minimum p-value rule, first suggested by Tippett (1931) and Wilkinson (1951), and advocated in Westfall and Young (1993). The intuition here is that the null hypothesis should be rejected if at least one of the individual t-statistics is sufficiently large in absolute value. For further discussion and other examples of test combination techniques, see Folks (1984), Dufour et al. (2015), and Catani and Ahlgren (2017).

In the present quantile regression context, it seems natural to assume that v_t in (4) has median zero and to estimate ϕ by least absolute deviations (LAD).[4] The LAD estimates of $\boldsymbol{\vartheta} = (\mu, \phi)'$ are found as

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} \sum_{t=1}^{T} \rho_\tau(x_t - \boldsymbol{\vartheta}' \boldsymbol{x}_{t-1})$$

by setting $\tau = 0.5$. In this case, the loss function corresponds to $\rho_{0.5}(v) = 0.5|v|$. The LAD

[4] Alternatively, it would be reasonable to proceed with OLS if we were willing to assume that v_t has mean zero and a finite variance.

estimates yield a t-ratio to test the value ϕ_0 specified in $H_0(\phi_0)$ taking the form

$$t(\phi_0) = \frac{\hat{\phi} - \phi_0}{\sqrt{\widehat{\text{var}}(\hat{\phi})}}, \tag{10}$$

whose computation is similar to (7). We let $\mathcal{T}(\phi_0)$ denote $|t(\phi_0)|$ with a view towards two-sided alternatives.

Let $\boldsymbol{z}_t = (y_t, x_t)'$ and define $\boldsymbol{\varepsilon}_t = (y_t,\ x_t - \phi_0 x_{t-1})'$, for $t = 1, ..., T$, given the value ϕ_0 in (8). Observe that $\boldsymbol{\varepsilon}_t = (\beta_0, \mu)' + \boldsymbol{\eta}_t$ and $\boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_T$ forms a collection of exchangeable random vectors under $H_0(\phi_0)$. This exchangeability property obviously holds no matter what the true values of β_0 and μ are, and no matter what the bivariate distribution of u_t and v_t is. To simplify the notation let $\tilde{\boldsymbol{\varepsilon}}_t = \boldsymbol{\varepsilon}_{d_t}$, where $d_1, ..., d_T$ is a permutation of the integers $1, ..., T$, and consider the sequence of random vectors $\tilde{\boldsymbol{z}}_1, ..., \tilde{\boldsymbol{z}}_T$ obtained from the recursion

$$\tilde{\boldsymbol{z}}_t = \begin{pmatrix} \tilde{y}_t \\ \tilde{x}_t \end{pmatrix} = \begin{pmatrix} 0 \\ \phi_0 \tilde{x}_{t-1} \end{pmatrix} + \tilde{\boldsymbol{\varepsilon}}_t, \text{ for } t = 1, ..., T, \tag{11}$$

with initial value $\tilde{\boldsymbol{z}}_0 = (0, x_0)'$. It is easy to see that $\boldsymbol{z}_t$ and $\tilde{\boldsymbol{z}}_t$ obey the same data-generating process when $H_0(\phi_0)$ holds.

Define $\boldsymbol{z}_t^{h_i} = (\sum_{k=0}^{h_i-1} y_{t+k}, x_t)'$, $\boldsymbol{Z}^{h_i} = [\boldsymbol{z}_1^{h_i}, ..., \boldsymbol{z}_{T-h_i+1}^{h_i}]$, and $\boldsymbol{Z} = [\boldsymbol{Z}^{h_1}, ..., \boldsymbol{Z}^{h_p}]$ from the original sample. An artificial sample generated according to (11) yields the analogues $\tilde{\boldsymbol{z}}_t^{h_i}$, $\widetilde{\boldsymbol{Z}}^{h_i}$, and $\widetilde{\boldsymbol{Z}}$. We then have under $H_0(\phi_0)$ that $\boldsymbol{Z} \overset{d}{=} \widetilde{\boldsymbol{Z}}$, for each of the $T!$ possible matrix realizations of $\widetilde{\boldsymbol{Z}}$. Theorem 1.3.7 in Randles and Wolfe (1979) further implies that

$$\begin{pmatrix} \mathcal{S} \\ \mathcal{T}(\phi_0) \end{pmatrix} \overset{d}{=} \begin{pmatrix} \widetilde{\mathcal{S}} \\ \widetilde{\mathcal{T}}(\phi_0) \end{pmatrix}, \tag{12}$$

where $\widetilde{\mathcal{S}} = \max_{1 \le i \le p} \max_{1 \le j \le q} |\tilde{t}(h_i, \tau_j)|$ and $\widetilde{\mathcal{T}}(\phi_0)$ refers to $|\tilde{t}(\phi_0)|$. Here $\tilde{t}(h_i, \tau_j)$ and $\tilde{t}(\phi_0)$ are the t-statistics in (7) and (10), respectively, each computed with the artificial data composing $\widetilde{\boldsymbol{Z}}$.

We see that generating such artificial samples according to (11) yields $T!$ possible vector values of $\big(\widetilde{\mathcal{S}}, \widetilde{\mathcal{T}}(\phi_0)\big)'$ which are all equally likely values for $\big(\mathcal{S}, \mathcal{T}(\phi_0)\big)'$, *i.e.*,

$$\Pr\left(\begin{pmatrix} \mathcal{S} \\ \mathcal{T}(\phi_0) \end{pmatrix} = \begin{pmatrix} \widetilde{\mathcal{S}} \\ \widetilde{\mathcal{T}}(\phi_0) \end{pmatrix}\right) = \frac{1}{T!}. \tag{13}$$

Based on this equally likely property, an α-level permutation test of, say, the null hypothesis that $\phi = \phi_0$ could be performed in the following way. Since large values of $\mathcal{T}(\phi_0)$ are more likely when $\phi \neq \phi_0$, the posited value ϕ_0 is rejected if the observed value of the test statistic $\mathcal{T}(\phi_0)$ falls in a set containing the $T!\alpha$ largest values of $\widetilde{\mathcal{T}}(\phi_0)$ that can be obtained from the class of all permutations.

Calculating the complete permutation distribution requires the careful enumeration of all $T!$ possibilities, which is well-nigh impossible with typical sample sizes. The computational burden of finding tail probabilities can be reduced by generating random samples according to (11) and computing $\big(\widetilde{\mathcal{S}}, \widetilde{\mathcal{T}}(\phi_0)\big)'$ each time. The relative frequencies of these values comprise the simulated permutation distribution. A very practical method to obtain the desired significance level and precise p-values, without performing a large number of random draws, is the MC test procedure first proposed by Dwass (1957). The procedure goes as follows. First, the statistics $\big(\mathcal{S}, \mathcal{T}(\phi_0)\big)'$ are computed with the original sample $\boldsymbol{Z}$. The test then proceeds by generating $B-1$ artificial samples $\tilde{\boldsymbol{Z}}_1, ..., \tilde{\boldsymbol{Z}}_{B-1}$, each one according to (11). With each such sample the statistics are computed yielding $\big(\widetilde{\mathcal{S}}_b, \widetilde{\mathcal{T}}_b(\phi_0)\big)'$, for $b = 1, ..., B-1$.

Including the observed values $\left(\mathcal{S}, \mathcal{T}(\phi_0)\right)'$, it is easy to see that the vectors $\left(\widetilde{\mathcal{S}}_1, \widetilde{\mathcal{T}}_1(\phi_0)\right)', ..., \left(\widetilde{\mathcal{S}}_{B-1}, \widetilde{\mathcal{T}}_{B-1}(\phi_0)\right)', \left(\mathcal{S}, \mathcal{T}(\phi_0)\right)'$ constitute B exchangeable draws from the bivariate uniform distribution in (13). To understand the idea of an MC test suppose αB is an integer, where α is the desired test level, and imagine sorting the B values $\widetilde{\mathcal{S}}_1, ..., \widetilde{\mathcal{S}}_{B-1}, \mathcal{S}$ in ascending order. If $\mathcal{S}$ ranks among the largest αB statistics, we would reject the null. In the present context, this happens with probability α under the null hypothesis. Since the permutation distribution is discrete, however, it is theoretically possible for ties to occur among the draws and this would cause an ambiguity when ranking the statistics.

Following Dufour (2006), we deal with that possibility by working with lexicographic ranks. These are obtained by first drawing B i.i.d. variates U_b, $b = 1, ..., B$, from a continuous uniform distribution on $[0, 1]$ and randomly assigning them to create the triplets $\left(\widetilde{\mathcal{S}}_1, \widetilde{\mathcal{T}}_1(\phi_0), U_1\right)', ..., \left(\widetilde{\mathcal{S}}_{B-1}, \widetilde{\mathcal{T}}_{B-1}(\phi_0), U_{B-1}\right)', \left(\mathcal{S}, \mathcal{T}(\phi_0), U_B\right)'$. From these, the tie-breaking ranks of $\mathcal{S}$ and $\mathcal{T}(\phi_0)$ are then computed as

$$\begin{aligned} \widetilde{R}_B\left[\mathcal{S}\right] &= 1 + \sum_{b=1}^{B-1} \mathbb{I}\left[\mathcal{S} > \widetilde{\mathcal{S}}_b\right] + \sum_{b=1}^{B-1} \mathbb{I}\left[\mathcal{S} = \widetilde{\mathcal{S}}_b\right] \times \mathbb{I}\left[U_B > U_b\right], \\ \widetilde{R}_B\left[\mathcal{T}(\phi_0)\right] &= 1 + \sum_{b=1}^{B-1} \mathbb{I}\left[\mathcal{T}(\phi_0) > \widetilde{\mathcal{T}}_b(\phi_0)\right] + \sum_{b=1}^{B-1} \mathbb{I}\left[\mathcal{T}(\phi_0) = \widetilde{\mathcal{T}}_b(\phi_0)\right] \times \mathbb{I}\left[U_B > U_b\right]. \end{aligned} \tag{14}$$

The key property (Dufour, 2006, Lemma 2.3) used to derive the MC p-values is that

$$\Pr\left(\widetilde{R}_B\left[\mathcal{S}\right] = b\right) = 1/B \text{ and } \Pr\left(\widetilde{R}_B\left[\mathcal{T}(\phi_0)\right] = b\right) = 1/B,$$

for $b = 1, ..., B$, when $H_0(\phi_0)$ holds. This simply says that the lexicographic ranks in (14) of exchangeable random variables are uniformly distributed over the integers $1, ..., B$.

From the marginal distributions of each statistic, we can then compute MC p-values as

$$\tilde{p}_B[\mathcal{S}] = \frac{B - \widetilde{R}_B[\mathcal{S}] + 1}{B},$$
$$\tilde{p}_B[\mathcal{T}(\phi_0)] = \frac{B - \widetilde{R}_B[\mathcal{T}(\phi_0)] + 1}{B}, \tag{15}$$

where $\widetilde{R}_B[\mathcal{S}]$ and $\widetilde{R}_B[\mathcal{T}(\phi_0)]$ are the ranks of $\mathcal{S}$ and $\mathcal{T}(\phi_0)$, respectively, given by (14). These p-values have the usual interpretation: $\tilde{p}_B[\mathcal{S}]$, for instance, is simply the proportion of $\widetilde{\mathcal{S}}$ values as extreme or more extreme than the observed $\mathcal{S}$ value in the simulated permutation distribution. Note that the two p-values in (15) may have a very complex dependence structure. Nevertheless, if we choose B so that αB is an integer (for $0 < \alpha < 1$), then these MC p-values each exactly have a size equal to α in the sense that

$$\Pr\Big(\tilde{p}_B[\mathcal{S}] \leq \alpha\Big) = \alpha,$$
$$\Pr\Big(\tilde{p}_B[\mathcal{T}(\phi_0)] \leq \alpha\Big) = \alpha, \tag{16}$$

under $H_0(\phi_0)$. See Ernst (2004) for more on the use of permutation methods for exact inference, and Dufour and Khalaf (2001) and Kiviet (2011, Ch. 6) for a general overview of MC test techniques and further references.

A decision rule could then be built from the logical equivalence that $H_0(\phi_0)$ is false if and only if $\{\bigcup_{i=1}^{p}\bigcup_{j=1}^{q}\beta_1(h_i,\tau_j) \neq 0\}$ or $\{\phi \neq \phi_0\}$. The critical region corresponding to this decision rule is $\{\tilde{p}_B[\mathcal{S}] \leq \alpha\} \cup \{\tilde{p}_B[\mathcal{T}(\phi_0)] \leq \alpha\}$ and, by subadditivity, we obtain its level as

$$\Pr\Big(\big\{\tilde{p}_B[\mathcal{S}] \leq \alpha\big\} \cup \big\{\tilde{p}_B[\mathcal{T}(\phi_0)] \leq \alpha\big\}\Big) \leq 2\alpha.$$

The marginal distributions of the p-values characterized by (16) are the foundations for the tests of H_0, developed next.

3.2 Tests of H_0

The expression in (9) makes clear that ϕ is a nuisance parameter in the present context, since it is not pinned down to a specific value under H_0. In order to test such a hypothesis, which contains several distributions, we can appeal to a *minimax* argument stated as: "reject the null hypothesis whenever, for all admissible values of the nuisance parameter under the null, the corresponding point null hypothesis is rejected" (Savin, 1984).

With the statistic $\mathcal{S} = \mathcal{S}(\phi_0)$,[5] this would mean maximizing the MC p-value $\tilde{p}_B[\mathcal{S}(\phi_0)]$ over $\phi_0 \in \mathcal{D}_\phi$. The rationale is that

$$\sup_{\phi_0 \in \mathcal{D}_\phi} \tilde{p}_B[\mathcal{S}(\phi_0)] \leq \alpha \Longrightarrow \tilde{p}_B[\mathcal{S}(\phi)] \leq \alpha,$$

where $\tilde{p}_B[\mathcal{S}(\phi)]$ is the MC p-value of $\mathcal{S}$ based on the true value ϕ. Moreover, $\Pr\big(\tilde{p}_B[\mathcal{S}(\phi_0)] \leq \alpha\big) = \alpha$ under $H_0(\phi_0)$ and for all $\phi_0 \in \mathcal{D}_\phi$. So, if αB is an integer, we then have

$$\Pr\left(\sup_{\phi_0 \in \mathcal{D}_\phi} \tilde{p}_B[\mathcal{S}(\phi_0)] \leq \alpha\right) \leq \alpha.$$

The decision rule in this case would be to reject H_0 if the maximized p-value is $\leq \alpha$. Otherwise, accept H_0 since there is not enough evidence against it. Note that this test has *level* α, meaning it is conservative.[6]

[5]From this point on, we shall use the notation $\mathcal{S}(\phi_0)$ to emphasize the fact that the distribution of $\mathcal{S}$ depends on ϕ_0.

[6]Here we follow the terminology in Lehmann and Romano (2005, Ch. 3) and say that a test of H_0 has *size* α if $\Pr(\text{ Rejecting } H_0 \mid H_0 \text{ true }) = \alpha$, and that it has *level* α if $\Pr(\text{ Rejecting } H_0 \mid H_0 \text{ true }) \leq \alpha$.

Following Beaulieu et al. (2007), we can replace $\mathcal{D}_\phi$ appearing in (9) by an exact confidence set for ϕ which is valid at least under the null hypothesis. This can be interpreted as plugging in an estimator of the (perhaps unknown) set of admissible ϕ-values. Let $C_\phi(\alpha_1)$ denote a confidence set for ϕ with level $1-\alpha_1$, *i.e.*, such that $\Pr\big(\phi \in C_\phi(\alpha_1)\big) \geq 1-\alpha_1$ under H_0. It can then be shown (see the Appendix) that

$$\Pr\left(\sup_{\phi_0 \in C_\phi(\alpha_1)} \tilde{p}_B\left[\mathcal{S}(\phi_0)\right] \leq \alpha_2\right) \leq \alpha_1 + \alpha_2. \tag{17}$$

Note as well that this is the main idea of the Bonferroni methods frequently used to deal with nuisance parameters in predictive mean regressions; see, for example, Cavanagh et al. (1995) and Campbell and Yogo (2006).[7]

Recall that a confidence set for a scalar parameter can be interpreted as the result of a collection of tests for each admissible value of the parameter. The confidence set simply reports all the values that cannot be rejected at a given nominal level. The confidence set appearing in (17) can therefore be obtained as

$$C_\phi(\alpha_1) = \big\{\phi_0 : \phi_0 \in \mathcal{D}_\phi,\ \tilde{p}_B\left[\mathcal{T}(\phi_0)\right] > \alpha_1\big\},$$

where $\tilde{p}_B\left[\mathcal{T}(\phi_0)\right]$ is the MC p-value of $\mathcal{T}(\phi_0)$ in (15). Observe that $C_\phi(\alpha_1)$ is an exact and distribution-free confidence set for ϕ, provided of course that $\mathcal{D}_\phi$ does not exclude the true

[7] It is worth remarking, however, that Phillips (2014) has shown that the confidence intervals based on local-to-unity limit theory used in these predictive regression tests are invalid in the stationary case, with zero asymptotic coverage probability. In particular, this causes the popular Q-test statistic of Campbell and Yogo (2006) to erroneously indicate predictability with probability approaching unity even though the null of no predictability holds true.

value of ϕ.[8] A confidence *interval* can be defined as

$$CI_\phi(\alpha_1) = \Big[\inf\big\{\phi_0 : \phi_0 \in C_\phi(\alpha_1)\big\},\ \sup\big\{\phi_0 : \phi_0 \in C_\phi(\alpha_1)\big\}\Big], \tag{18}$$

which may be used to conduct finite-sample inference about the value of ϕ, since $\Pr\big(\phi \in CI_\phi(\alpha_1)\big) \geq 1 - \alpha_1$. This is an interesting result in itself because it is well known that the conventional t-test can yield misleading conclusions about the AR(1) parameter, particularly when it is close to 1; see Tanaka (1983), Nankervis and Savin (1985, 1988), Rayner (1990), and Nankervis and Savin (1996).

3.3 Summary of MMC test procedure

The test of H_0 in (9) proceeds according to the following steps:

1. For each $\phi_0 \in \mathcal{D}_\phi$, compute the MC p-values $\big(\tilde{p}_B\left[\mathcal{S}(\phi_0)\right], \tilde{p}_B\left[\mathcal{T}(\phi_0)\right]\big)$ as in (15).

2. Compute a confidence set for ϕ with level $1 - \alpha_1$ as

$$C_\phi(\alpha_1) = \big\{\phi_0 : \phi_0 \in \mathcal{D}_\phi,\ \tilde{p}_B\left[\mathcal{T}(\phi_0)\right] > \alpha_1\big\}.$$

3. Find $p^* = \sup_{\phi_0 \in C_\phi(\alpha_1)} \tilde{p}_B\left[\mathcal{S}(\phi_0)\right]$ and reject H_0 if $p^* \leq \alpha_2$. Otherwise, accept H_0.

In practical applications, the set $\mathcal{D}_\phi$ in Step 1 is replaced by a discrete grid. Observe that only the candidate values ϕ_0 vary in Step 1. In order to control the underlying randomness in the computation of the MC p-values linked to each ϕ_0-value, the seed of the random number generator should be reset to the same value before each considered ϕ_0. In the terminology

[8] On the other hand, eliminating (truncating) inadmissible values from a confidence set does not modify its level (Abdelkhalek and Dufour, 1998).

of Dufour (2006), we will refer to the test described by Steps 1–3 as a *maximized* MC (MMC) test. As a by-product, this MMC test procedure also yields an exact distribution-free confidence set for ϕ in Step 2. A confidence interval for ϕ can then be obtained from $C_\phi(\alpha_1)$ according to (18).

Given a desired significance level α, we see there is a tradeoff between the width of the confidence set $C_\phi(\alpha_1)$ in Step 2 and the significance level $\alpha_2 = \alpha - \alpha_1$ in Step 3. While the choice of α_1, α_2 has no effect on the overall level (as long as $\alpha_1 + \alpha_2 = \alpha$), it does matter for power. Campbell and Dufour (1997) suggest that it is better to take a wider confidence set in order to have a tighter critical value when deciding whether to reject H_0. Accordingly, we carry on with the testing strategy represented by $\alpha_1 = 1\%, \alpha_2 = 4\%$ for an overall $\alpha = 5\%$.

A computationally simplified procedure is obtained by replacing the confidence set by the point estimate $\hat{\phi}$ and rejecting H_0 whenever $\tilde{p}_B\big[\mathcal{S}(\hat{\phi})\big] \leq \alpha$. Dufour (2006) refers to such tests as *local* MC (LMC) tests, and, since they are quite natural, we will study their size and power properties in the simulation study.[9] These LMC tests are also a very natural benchmark against which to compare the theoretically sound MMC tests.

4 Simulation Experiments

In this section, we report the results of a series of simulation experiments designed to examine the size and power performance of the proposed tests. All the tests are performed at the nominal $\alpha = 5\%$ significance level with $B-1 = 99$ artificial samples, and the reported results are based on 1000 replications of each data-generating configuration.

[9]The term "local" reflects the fact that the underlying MC p-value is based on a specific choice for the nuisance parameter. Under additional regularity conditions, Dufour (2006) shows the asymptotic validity of LMC tests. These conditions are notably more restrictive than those under which (17) obtains.

4.1 Design and results

We generate data according to

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_{t-1} + \sigma_t u_t, \\ x_t &= \mu + \phi x_{t-1} + v_t, \end{aligned} \tag{19}$$

where $u_t = \gamma v_t + \sqrt{1-\gamma^2}\xi_t$ with $\gamma \in [-1, 1]$. Here v_t and ξ_t are independent with $v_t \sim N(0, 1)$ and $\xi_t \sim t(3)$. In this setup we have $\text{Corr}(u_t, v_t) = \gamma/\sqrt{3 - 2\gamma^2}$, so the parameter γ determines the strength of feedback from u_t to future values of the regressor variable. Observe also that the elements within $\boldsymbol{\eta}_t = (u_t, v_t)'$ are not exchangeable. We set $\beta_0 = \mu = 0$, and let $\gamma = 0$, -0.95 to examine the effects of feedback.[10] We present results for the sample size $T = 240$ and vary the persistence of x_t according to $\phi = 0.5$, 0.95, 0.99, 1, and 1.01.

The null hypothesis H_0 is represented by $\beta_1 = 0$ and $\sigma_t = 1$, leaving no way for x_{t-1} to affect the distribution of y_t. We let the alternative hypothesis be parameterized as $\beta_1 = 0.1$ and $\sigma_t = |x_{t-1}|$ so the τth conditional quantile of y_t becomes

$$Q_\tau(y_t \,|\, x_{t-1}) = \beta_0 + \beta_1 x_{t-1} + |x_{t-1}| Q_\tau(u_t),$$

where $Q_\tau(u_t)$ is the τth quantile of u_t. This specification bears a resemblance to the absolute value ARCH model of Taylor (1986) and Schwert (1989) in which the conditional standard deviation depends on the absolute values of the volatility forcing variables. Since u_t is symmetrically distributed about zero, we have $Q_\tau(u_t) = 0$ at $\tau = 0.5$ (the median). This means that the predictive ability of x_{t-1} equals $\beta_1 x_{t-1}$ for the conditional median of y_t and

[10]The value $\gamma = -0.95$ implies $\text{Corr}(u_t, v_t) = -0.87$, which is in the range of error correlations typically found in the stock return predictability literature.

increases (symmetrically) for the outer quantiles of y_t, because $|x_{t-1}||Q_\tau(u_t)|$ is an increasing function of $|\tau - 0.5|$. This setup is motivated by the empirical evidence in Cenesizoglu and Timmermann (2008), Maynard et al. (2011), and Lee (2016) who suggest that many economic variables have far greater predictive ability for the outer quantiles rather than for the centre of the return distribution.

For comparison purposes, we report the size and power of the standard t-test and the IVX-QR test of Lee (2016). The IVX-QR method proceeds by generating an instrumental variable w_t according to the filtering rule

$$w_t = R_w w_{t-1} + \Delta x_t, \quad R_w = 1 + \frac{c_w}{T^\delta},$$

where $w_0 = 0$, and $\delta \in (0,1)$ and $c_w < 0$ are user-chosen parameters. In order to test the null of no quantile predictability at $h = 1$, Lee (2016) proposes to use an ordinary quantile regression of y_t on w_{t-1}, and shows that the resulting estimator is asymptotically normal. Here we set $c_w = -5$ as in Lee (2016) and report the test results for a range of δ-values to investigate the effects of that parameter.

Table 1 show the empirical size of the standard t- and IVX-QR tests for predictability at horizon $h = 1$, applied at various quantile levels from 0.1 to 0.9, one at a time. The t-test behaves reasonably well when there is no feedback ($\gamma = 0$), except at the extreme outer quantile levels $\tau = 0.1$, 0.9, where it tends to be more oversized. In the presence of feedback ($\gamma = -0.95$), however, the t-test shows a tendency to over-reject at all quantile levels, especially as ϕ increases. The behaviour of the IVX-QR test also depends very much on the strength of feedback. When $\gamma = 0$, the IVX-QR test behaves essentially like the

t-test, no matter the choice of δ. An exception occurs when $\delta = 0.1$, in which case IVX-QR over-rejects far more for the central quantiles. On the flip side, the choice of δ matters a great deal when $\gamma = -0.95$. In that case, a lower value of δ seems preferable to keep the test size closer to the nominal value. We see however that the asymptotic test has more difficulty controlling size at the extreme outer quantiles, where there are relatively fewer observations.

Tables 2 and 3 report the size of the LMC and MMC tests across quantiles $\tau = 0.1, ..., 0.9$ and horizons $h = 1, 2, 3$. For the most part, the LMC tests appear quite well behaved with empirical size close to 5%. The LMC approach starts running into trouble when $\gamma = -0.95$ and $\phi \geq 1$. In those cases, we see LMC test sizes as high as 15%. Interestingly, the LMC tests behave relatively better for tail quantiles compared to the more central quantiles. In accordance with the developed theory, the MMC tests are seen in Table 3 to always respect the nominal 5% level constraint. This holds true regardless of the predictor persistence and feedback strength, and whether the test is performed at individual quantile levels τ or individual horizons h, or for multiple quantile levels (all τ) and/or multiple horizons (all h).

Table 4 shows the size-adjusted power of the standard t-test and the IVX-QR test. We can see that the t-test generally achieves better power. The performance of the IVX-QR test tends to improve as δ is set higher. For example, $\delta = 0.9$ tends to deliver better power than when δ is set to lower values. An exception seems to occur around $\tau = 0.5$. Indeed for the central quantiles, setting $\delta = 0.1$ tends to yield better power, especially as ϕ increases to 0.99 and beyond. In fact, when ϕ is in that range, the power of IVX-QR ($\delta = 0.9$) follows a U-shaped pattern: lowest for the central quantiles and generally increasing towards the outer quantiles. Whereas when ϕ decreases, the power pattern gets inverted: higher around the central quantiles and decreasing towards the outer quantiles.

Tables 5 and 6 report the power of the LMC and MMC tests. The power calculations for the LMC tests are based on size adjustments. The MMC tests did not require any such adjustments since they reject the null with probability not exceeding 5%. We see that the power of LMC ($h = 1$) is not far behind that of the t-test in Table 4, particularly as the predictor persistence increases. A striking result seen in Tables 5 and 6 is that the LMC and MMC procedures can deliver more power when testing for predictability at multiple quantiles (all τ), particularly as the predictor becomes more persistent. Indeed, as ϕ increases we see the power of the 'all τ' tests exceeding the power attained at any of the individual τ-values.

A comparison of Tables 5 and 6 shows that the LMC tests tend to be more powerful than their MMC counterparts. However, it is also important to bear in mind that size-adjusted tests are not feasible in practice, but merely serve here as a benchmark for the truly exact MMC tests. Note also from these tables that the LMC-MMC power gap closes as the predictor persistence increases and feedback strengthens. Finally, we can observe that power does not generally increase with the prediction horizon, h. The multiple-horizons 'all h' tests can nonetheless be more powerful than the specific-horizon tests. We see this occurring in Tables 5 and 6 for the more central quantiles, notably when $\gamma = -0.95$ (presence of feedback) and $\phi \geq 0.95$ (persistent predictor).

4.2 Robustness assessment

We next assess the robustness to departures from the assumptions maintained under (4). First, we examine what happens to the empirical size of the proposed LMC and MMC tests when the predictor variable follows an AR(2) process of the form:

$$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + v_t,$$

with $\phi_1 = 1.30$ and $\phi_2 = -0.31$. Second, we investigate the size performance in the presence of GARCH effects appearing as

$$\begin{aligned} y_t &= \beta_0 + \sigma_t u_t, \\ \sigma_t^2 &= \omega + a_1(y_{t-1} - \beta_0)^2 + b_1 \sigma_{t-1}^2, \end{aligned}$$

where $\omega = 0.01$, $a_1 = 0.10$, and $b_1 = 0.85$. The parameters values chosen here are in line with what is found empirically with the monthly data used in Section 5.

Table 7 shows the empirical size in percentage of the proposed tests under: (i) the AR(2) process with i.i.d. errors; (ii) the GARCH(1,1) process along side a correctly assumed AR(1) model (with ϕ set to 0.99); and (iii) the AR(2) and GARCH(1,1) processes together. In each case, the feedback parameter is set as $\gamma = -0.95$. Comparing the LMC and MMC tests, we see immediately that the LMC tests are far more sensitive to departures from the maintained assumptions. In particular, GARCH errors are quite damaging for the empirical size of the LMC tests. On the contrary, the MMC tests, owing to their conservativeness, offer more robustness to violations of the assumptions.

4.3 Sign-based predictability tests

For testing predictability at $\tau = 0.5$ and $h = 1$, Campbell and Dufour (1997) introduce the following sign-based statistics:

$$S(b) = \sum_{t=1}^{T} s\big[(y_t - b)g_{t-1}\big],$$
$$W(b) = \sum_{t=1}^{T} s\big[(y_t - b)g_{t-1}\big]R_t^+(b),$$

where $s[z] = 1$ when $z \geq 0$, and $s[z] = 0$ when $z < 0$; $R_t^+(b)$ is the rank of $|y_t - b|$ when $|y_1 - b|, ..., |y_T - b|$ are placed in ascending order; and $g_t = x_t - \hat{m}_t$, with $\hat{m}_t = \text{median}\{x_0, ..., x_t\}$. These sign and signed rank statistics yield median predictability tests that allow for GARCH effects in y_t, and remain exact without any modelling assumptions whatsoever for x_t.

Under the null hypothesis of no median predictability $\beta_1 = 0$, $S(\beta_0)$ follows a binomial distribution $\text{Bi}(T, 1/2)$ and $W(\beta_0)$ is distributed like the Wilcoxon variate. Let $S = \inf\{|S^*(b)| : b \in CI_{\beta_0}(\alpha_1)\}$, where $S^*(b) = (S(b) - T/2)/\sqrt{T/4}$. If we adopt the two-sided decision rule: *reject the null of no median predictability when* S *is significant at level* $\alpha_2/2$, then the overall level of this two-step procedure is bounded by $\alpha = \alpha_1 + \alpha_2$ (Campbell and Dufour, 1997, Proposition 2). The same argument applies to $W = \inf\{|W^*(b)| : b \in CI_{\beta_0}(\alpha_1)\}$ with $W^*(b) = (W(b) - T(T+1)/4)/\sqrt{T(T+1)(2T+1)/24}$.

The first-step confidence interval $CI_{\beta_0}(\alpha_1)$ appearing in the definition of S is obtained by inverting the exact sign test $\mathcal{B} = \sum_{t=1}^{T} s\big[(y_t - b)\big]$, which follows $\text{Bi}(T, 1/2)$ when $b = \beta_0$. For the W statistic, $CI_{\beta_0}(\alpha_1)$ is obtained by inverting a Wilcoxon signed rank test $\mathcal{W} = \sum_{t=1}^{T} s\big[(y_t - b)\big]R_t^+(b)$ for the median. It also seems natural to replace the unknown β_0 by an estimate $\hat{\beta}_0 = \text{median}\{y_1, ..., y_T\}$. We denote these median-estimate tests by $S(\hat{\beta}_0)$ and $W(\hat{\beta}_0)$ and examine their performance in the simulation experiments. In order to make the experiments comparable to Campbell and Dufour (1997), we set $\sigma_t = 1$ in (19) and focus on median predictability by letting $\beta_1 = 0$ under the null, and then $\beta_1 = 0.05$, 0.1 under the alternative hypothesis.

We observe again in Table 8 the size of the LMC test tending to deviate from the nominal 5% level when $\gamma = -0.95$ and ϕ increases beyond 0.99. The median-estimate $S(\hat{\beta}_0)$ and $W(\hat{\beta}_0)$ tests appear better behaved. The size of the MMC test and the sign-based S and W tests stays below the nominal level, as expected. The power results in Table 8 for the LMC, $S(\hat{\beta}_0)$ and $W(\hat{\beta}_0)$ tests are based on size-adjustments; no such adjustments are needed for the MMC, S and W tests. We see clearly the LMC test dominating the power ranking. As ϕ increases, the MMC test overtakes the suite of sign-based tests. However, with a less persistent predictor, the $S(\hat{\beta}_0)$ and $W(\hat{\beta}_0)$ do better than the MMC test. We can also see the S and W tests losing some of their power when ϕ increases to 1 and 1.01.

5 Finite-Sample Return Predictability

5.1 Data description

In this section, we further illustrate the new quantile predictability tests with an application to the monthly and weekly excess returns (r) on the S&P value-weighted stock market index from January 1962 to December 2015 (648 monthly, 2817 weekly observations). We consider several return horizons ($h = 1, 3, 12, 60, 120$ months; and $h = 1, 4, 12, 24, 52$ weeks) and the six predictors we employ are commonly used in the stock return predictability literature:

log dividend-price ratio (*d/p*), log earnings-price ratio (*e/p*), book-to-market ratio (*btm*), default yield spread (*dfy*), term spread (*tms*), and short-term interest rate (*tbl*). The first three predictors (*d/p*, *e/p*, *btm*) are valuation ratios based on stock characteristics, while the last three (*dfy*, *tms*, *tbl*) are related to interest rates. These data are a subset of those used by Goyal and Welch (2008) and all the monthly series were obtained directly from Amit Goyal's website.

We constructed the weekly data set following the methodology in Goyal and Welch (2008). To compute weekly excess returns on the S&P 500 index, we downloaded the daily adjusted closing prices from Yahoo Finance and computed the end-of-week return values as the continuously compounded daily returns from Monday to Friday. The weekly excess returns were then obtained by subtracting the 3-month Treasury bill rate, obtained from the economic research database of the Federal Reserve Bank of St. Louis (FRED).

Similar to Goyal and Welch (2008), the weekly valuation ratios are based on data from Robert Shiller's website. To compute weekly dividends (and earnings), we took 52-week moving sums of dividends paid (earnings, respectively) on the S&P 500 index. Book values from 1962 to 2005 are from Value Line's *Long Term Perspective Chart*. For the period from 2006 to 2015, we extracted the book values using Goyal and Welch's monthly data. We first multiplied their monthly book-to-market ratio with the end-of-month market value of the Dow Jones Industrial Average (DJIA). Then we computed the weekly book-to-market as the ratio of book value and end-of-week market price of the DJIA. Following Goyal and Welch (2008), for the weeks in March to December, *btm* is computed as the ratio of book value at the end of the previous year and the price at the end of the current week. For the weeks in January to February, *btm* is the book value at the end of two years ago divided by the price at the end of the current week. Finally, the predictors based on the interest rates are computed using weekly data from FRED.

5.2 Empirical findings

Table 9 reports some summary statistics for each variable, and the LAD-based estimates of the assumed AR(1) model for each of the six predictors in turn. Beneath the LAD point estimates $\hat{\phi}$ are reported in parentheses the associated standard errors $\hat{\sigma}(\hat{\phi})$ and in square brackets the $1 - \alpha_1 = 99\%$ confidence intervals for ϕ, obtained by inverting the MC t-ratio test according to (18). At the monthly (weekly) frequency the point estimates $\hat{\phi}$ are between 0.967 (0.996) and 1.003 (1.000), which suggests that the predictors are highly persistent. In fact, the (very tight) confidence intervals reveal that the unit root hypothesis cannot be rejected for any of the predictors. These results are consistent with those of Valkanov (2003), Torous et al. (2004), and Hjalmarsson (2011) who also report confidence intervals (by inverting unit root tests) for commonly used stock return predictors.

The unit root findings may be unsavoury from a theoretical point of view. Indeed, present value models which impose transversality imply that the dividend-price ratio must be stationary. Chief among these is the Campbell and Shiller (1988) present-value equation, which is based on a log-linear approximation of returns that relates log stock returns linearly to log prices and log dividends. That approach breaks down when the log dividend-price ratio has a unit root, because it then has no fixed mean around which to take a Taylor series approximation.[11] So we also ran our tests with the set $\mathcal{D}_\phi$ restricted to the open interval $(-1, 1)$. Excluding the possibility of unit roots did not change any of the test results

[11] Campbell (2008) proposes an alternative approach that is valid when the dividend-price ratio follows a geometric random walk.

presented next.

Tables 10 and 11 (12 and 13) report the p-values of the proposed LMC and MMC tests, respectively, with the monthly (weekly) data. They evaluate the predictability of monthly (weekly) excess stock returns over a range of quantile levels τ from 0.05 to 0.95 and prediction horizons h from 1 month (1 week) to 120 months (52 weeks). The last column (all τ) shows the p-value for the joint hypothesis of no return predictability at any of the considered quantile levels, given the horizon h for the corresponding line; the last row (all h) gives the p-value for the joint hypothesis that, at the given return quantile level τ in the corresponding column, there is no predictability at any of the considered horizons; and the entry at the intersection of 'all τ' and 'all h' is the p-value for the joint null of no predictability at any quantile level and any horizon. The entries in bold correspond to cases of statistical significance at the 5% level.

The LMC and MMC p-values reveal that the valuation ratios (*d/p*, *e/p*, *btm*) essentially have no predictive ability for stock return quantiles at any prediction horizon. This finding extends the conclusion reached by Hjalmarsson (2011) that these three valuation ratios have no predictive ability for the mean of the stock return distribution. We can see a few exceptions, like the book-to-market ratio (*btm*) in Table 11 for $\tau = 0.95$ when $h = 1$. However, the weekly results in Tables 12 and 13 make clear that those few instances are fragile to the choice of sampling frequency.

The pattern that emerges from Tables 10–13 is that any evidence of quantile predictability is mainly from predictors related to the interest rates. In particular, the default yield spread (*dfy*) appears predictive of the right tail and the short-term interest rate (*tbl*) appears predictive of the more central quantiles. Moreover, this predictability evidence is stronger for short prediction horizons (typically up to 3 months). The lines labeled 'all h' in Tables 11 and 13 have conservative p-values that are far from the 5% cutoff, indicating that none of the variables can predict stock return quantiles over all the considered prediction horizons.

Table 14 reports the sample (unconditional) quantiles $\hat{Q}_\tau(r_t)$ of the monthly excess returns. The LMC and MMC predictability test results suggest that the default yield and the short rate generate statistically significant slope coefficients in the predictive quantile regressions. The lines indicated by $\hat{\beta}_1(\tau)$ in Table 14 report those slope coefficients at horizon $h = 1$ month and the entries in bold are the instances where the MMC test result (in Table 11) is significant at the 5% level. This table also reports $R^2(\tau)$ which is the goodness-of-fit measure developed by Koenker and Machado (1999) for quantile regressions. Like the conventional OLS-based R^2, the $R^2(\tau)$ measure lies between 0 and 1, with higher values here indicating a better fit in the sense that the τth conditional quantile function $\hat{Q}_\tau(r_t|x_{t-1})$ "significantly" depends on x_{t-1}. As typically found in predictive mean regression models of stock returns,[12] the $R^2(\tau)$ values found here are also quite low. This raises the question of whether the quantile predicability is economically significant.

5.3 Economic significance

To explore this question, we compare how much the predicted conditional quantiles $\hat{Q}_\tau(r_t|x_{t-1})$ vary over time relative to their unconditional counterparts, $\hat{Q}_\tau(r_t)$.[13] For each predictor, Table 14 reports the ratio $\frac{\hat{\sigma}[\hat{Q}_\tau(r_t|x_{t-1})]}{|\hat{Q}_\tau(r_t)|}$ which is the estimated standard deviation of the conditional quantile relative to the absolute value of the unconditional quantile. We can

[12]Campbell (2018, p. 139) provides an explanation as to why the conventional R^2 is expected to be low when the predictor is governed by a persistent AR(1) model.

[13]This gauge of economic significance is based on Cochrane (2008) who examines the predictability of the dividend-price ratio for the mean return.

think of this quantity as an "economic" measure of how much the conditional quantiles vary over time. Table 14 reveals that the conditional quantiles vary a great deal. For instance the unconditional quantiles of excess returns at $\tau = 0.8, 0.9, 0.95$ are 3.78%, 5.37%, and 7.03%, respectively. The time variation of the corresponding conditional quantiles, given the default yield spread (*dfy*) as predictor, is 19.6%, 19.7%, and 16.9% as large as the respective unconditional quantile values. Even greater is the time variation by 203% (albeit of -0.27% at $\tau = 0.4$) and 60.6% (of 0.80% at $\tau = 0.5$) of the conditional excess return quantiles with the short-term interest rate (*tbl*) as predictor.

5.4 Discussion

How does the quantile predictability evidence uncovered here relate to the extant literature on mean return predictability? For the U.S. postwar data, mean stock return predictability has almost become a stylized fact. One of the central tenets of this literature is the Campbell and Shiller (1988) present-value equation, which says that deviations of the log dividend-price ratio from its steady-state level ought to predict either future log returns, or future log dividend growth rates, or both. A commonly held view is that the dividend-price ratio (*i.e.*, the dividend yield) does not predict dividend growth rates, but instead predicts returns and this predictability is stronger at longer horizons; see Campbell (1991) and Cochrane (1992, 2008), among others. A number of other financial and macroeconomic variables are often believed to also have the ability to predict future stock returns; see Van Nieuwerburgh and Koijen (2009) for a survey.

In Tables 10–13, the dividend-price ratio was found to have no predictive ability for stock returns. This is perhaps not surprising given that the Campbell-Shiller present-value equation is derived from a log-linear approximation of returns, which in turn relies on the stationarity of the dividend-price ratio. Indeed, the accuracy of the log-linear approximation depends on the variability of the log dividend-price ratio around its steady-state level (cf. Campbell, 2018, §5.3). From Table 9, the dividend-price ratio appears to be non-stationary, which suggests that the Campbell-Shiller approximation may be rather crude. Other authors (*e.g.* Goyal and Welch, 2003; Park, 2010; Dybvig and Zhang, 2018, among others) also find evidence of a unit root in the dividend-price ratio. Boudoukh et al. (2007), Robertson and Wright (2006), and Park (2010) attribute this I(1) behaviour of the dividend-price ratio to changes in corporate payout policy since the 1980s.[14] They argue that share repurchases instead of dividend payouts caused a break in the cointegrating relation between log prices and log dividends. Park (2010) shows that, when it is I(0), the dividend-price ratio may have predictive power for future stock returns, which are well known to be I(0). But when the dividend-price ratio is I(1), the standard predictive regression becomes unbalanced and the dividend-price ratio's OLS slope coefficient goes to zero.

More recently, Choi et al. (2016) also find that the dividend yield's predictive ability disappears once the peculiar characteristics of stock return predictive regressions are properly addressed. Indeed, even a simple predictive regression of one-period-ahead stock returns onto dividend yields is subject to many econometric issues. Since Mankiw and Shapiro (1986) and Stambaugh (1999), we know that the use of traditional asymptotic theory is problematic when there is feedback from returns to future values of the regressor and the regressor variable itself is highly persistent over time. Inference in long-horizon regressions is further complicated because the error terms contain a moving-average structure, induced

[14] Specifically, the enactment of SEC Rule 10b-18 in 1982 spawned an explosion in repurchase activity that had a profound effect on the manner in which firms distribute earnings to their shareholders (Boudoukh et al., 2007).

by summing returns over long horizons. Valkanov (2003) and Boudoukh et al. (2008) show that in the presence of a highly persistent regressor, predictability may artificially emerge in a standard regression as the prediction horizon increases. Moreover, "robust" standard errors (*e.g.* Hansen and Hodrick, 1980; Newey and West, 1987) tend to perform poorly in finite samples, since the residual serial correlation can be very pronounced due to the overlap in the data.

An emerging view is that, once the moving-average error structure is properly accounted for, the evidence of mean return predictability does not typically become stronger as the horizon increases; see Valkanov (2003), Torous et al. (2004), Ang and Bekaert (2007), Boudoukh et al. (2008), Hjalmarsson (2011), and Kostakis et al. (2015). In particular, Ang and Bekaert (2007) conclude that "the strongest [mean] predictability comes from the short rate rather than from the dividend yield" and that "predictability is mainly a short-horizon, not a long-horizon, phenomenon."[15]

Our empirical findings reveal an expanded view of this phenomenon. We too find that the short rate has predictive ability for the central quantiles, but also that the default yield spread predicts the right-tail quantiles of the return distribution. With our test procedure that allows for a highly persistent regressor and accounts for the overlapping observations used in long-horizon regressions, we find that the quantile predictability evidence using these interest rate variables is strongest at short horizons. In line with Ang and Bekaert (2007), Hjalmarsson (2011), and Kostakis et al. (2015) who focus on the mean, there is little to no evidence of predictability using the valuation ratios (dividend-price, earnings-price, book-to-market) for any part of the return distribution. To quote Ang and Bekaert (2007), "the

[15] Ang and Bekaert (2007) also develop a non-linear present-value model with stochastic discount rates, short rates, and dividend growth that rationalises the predictability findings.

predictability [by the dividend yield] that has been the focus of most recent finance research is simply not there."

6 Concluding Remarks

We have developed a procedure to test for long-horizon quantile predictability that allows for the presence of an endogenous and persistent predictor variable along with heavy-tailed errors in both the predictive quantile regression model and the AR(1) model, assumed for the predictor variable. As far as we know, our approach is the only one that allows multiple horizons and quantile levels to be tested jointly. We achieve this by exploiting the technique of MMC tests, developed in Dufour (2006). Specifically, we combine an exact confidence set for the persistence parameter of the AR(1) model with conditional distribution-free MC tests of no quantile predictability that are linked to each point in the confidence set.

The confidence set is obtained by "inverting" a distribution-free MC t-ratio test for the AR(1) parameter, and the predictability tests are based on a data-dependent combination of quantile regression t-statistics, computed at each considered horizon and quantile level. This approach yields MMC permutation tests of the joint null hypothesis of no quantile predictability whose familywise error rate (probability of committing at least one Type I error) is kept under control, no matter the sample size. To complement the results establishing the validity of our procedure, a simulation study made clear that the proposed tests deliver good discriminatory power in detecting quantile predictability. Finally, an empirical application revealed that certain parts of the distribution of stock returns are indeed predictable by variables related to interest rates.

These results beg the question: what is the economic value of the predictability findings?

In order to answer this question, consider an investor who allocates some wealth W_T to a portfolio comprising a risk-free asset and a risky asset in order to maxime the expected utility $E[U(W_{T+1})]$ over next period wealth. The solution to this problem involves the conditional distribution of next period's excess return on the risky asset. Following Cenesizoglu and Timmermann (2008), that distribution can be approximated using the out-of-sample quantile forecasts made in period T. The economic value of quantile predictability can then be evaluated through the certainty equivalent return or any other measure of portfolio performance. Jondeau and Rockinger (2006) use a Taylor series expansion to get an approximation of $E[U(W_{T+1})]$ that depends on the moments up to order four of the portfolio return distribution. Instead of the conventional measures of higher moments they use, an extension of the robust quantile-based measures in Kim and White (2004) to the conditional case would offer an alternative way to assess the value of quantile predicability. Such investigations of out-of-sample performance go beyond the scope of the present paper, so we leave them for future research.

A final remark is in order. The empirical analysis we presented in this paper is still subject to a multiple comparisons problem since the predictability tests were performed with each predictor variable separately. Indeed, the null hypothesis of no predictability implies that all predictors should be jointly insignificant. Extending our methodology to deal with this important issue is a challenge we intend to take up in future work.

Appendix

The proof closely follows that of Proposition 2 in Campbell and Dufour (1997). It is included here for completeness. We wish to show that $\Pr\left(\sup_{\phi_0 \in C_\phi(\alpha_1)} \tilde{p}_B\left[\mathcal{S}(\phi_0)\right] \leq \alpha_2\right) \leq \alpha_1 + \alpha_2$.

This will be true if $\Pr(A) \le \alpha_1 + \alpha_2$, where A is the event $\tilde{p}_B[\mathcal{S}(\phi_0)] \le \alpha_2$ for all $\phi_0 \in C_\phi(\alpha_1)$.

Define the set $I = \{\phi_0 : \phi_0 \in C_\phi(\alpha_1) \text{ and } \tilde{p}_B[\mathcal{S}(\phi_0)] > \alpha_2\}$. Then, via Bonferroni's inequality, we have that

$$\begin{aligned} \Pr(\phi \in I) &= 1 - \Pr\left(\phi \notin C_\phi(\alpha_1) \text{ or } \tilde{p}_B[\mathcal{S}(\phi_0)] \le \alpha_2]\right) \\ &\ge 1 - \Pr(\phi \notin C_\phi(\alpha_1)) - \Pr\left(\tilde{p}_B[\mathcal{S}(\phi_0)] \le \alpha_2]\right) \\ &\ge 1 - \alpha_1 - \alpha_2, \end{aligned}$$

since $\Pr\left(\phi \in C_\phi(\alpha_1)\right) \ge 1 - \alpha_1$ by definition of the confidence set for ϕ, and $\Pr\left(\tilde{p}_B[\mathcal{S}(\phi_0)] \le \alpha_2]\right) = \alpha_2$ from (16). Observe that $\Pr(A) = \Pr(B^c)$, where B is the event $\tilde{p}_B[\mathcal{S}(\phi_0)] > \alpha_2$ for some $\phi_0 \in C_\phi(\alpha_1)$. Note also that $\phi \in I \Longrightarrow B$. Hence

$$\Pr(B) \ge \Pr(\phi \in I) \ge 1 - \alpha_1 - \alpha_2,$$

which implies the desired result: $\Pr(A) \le \alpha_1 + \alpha_2$.

References

Abdelkhalek, T. and J.-M. Dufour (1998). Statistical inference for computable general equilibrium models, with application to a model of the Moroccan economy. *Review of Economics and Statistics 80*, 520–534.

Amihud, Y. and C. Hurvich (2004). Predictive regression: a reduced-bias estimation approach. *Journal of Financial and Quantitative Analysis 39*, 813–841.

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

Bandi, F. and B. Perron (2008). Long-run risk-return trade-offs. *Journal of Econometrics 143*, 349–374.

Bandi, F., B. Perron, A. Tamoni, and C. Tebaldi (2018). The scale of predictability. *Journal of Econometrics 208*, 120–140.

Barnard, G. (1963). Comment on 'The spectral analysis of point processes' by M.S. Bartlett. *Journal of the Royal Statistical Society (Series B) 25*, 294.

Bassett, G., R. Koenker, and G. Kordas (2004). Pessimistic portfolio allocation and Choquet expected utility. *Journal of Financial Econometrics 2*, 477–492.

Beaulieu, M.-C., J.-M. Dufour, and L. Khalaf (2007). Multivariate tests of mean-variance efficiency with possibly non-Gaussian errors. *Journal of Business and Economic Statistics 25*, 398–410.

Birnbaum, Z. (1974). Computers and unconventional test statistics. In F. Proschan and R. Serfling (Eds.), *Reliability and Biometry*, pp. 441–458. SIAM, Philadelphia.

Boudoukh, J., R. Michaely, M. Richardson, and M. Roberts (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *Journal of Finance 62*, 877–915.

Boudoukh, J. and M. Richardson (1993). Stock returns and inflation. *American Economic Review 64*, 1346–1355.

Boudoukh, J., M. Richardson, and R. Whitelaw (2008). The myth of long-horizon predictability. *Review of Financial Studies 21*, 1577–1605.

Campbell, B. and J.-M. Dufour (1997). Exact non-parametric tests of orthogonality and random walk in the presence of a drift parameter. *International Economic Review 38*, 151–173.

Campbell, J. (1991). A variance decomposition for stock returns. *Economic Journal 101*, 157–179.

Campbell, J. (2008). Viewpoint: Estimating the equity premium. *Canadian Journal of Economics 41*, 1–21.

Campbell, J. (2018). *Financial Decisions and Markets: A Course in Asset Pricing.* Princeton University Press, New Jersey.

Campbell, J. and R. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies 1*, 195–228.

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

Catani, P. and N. Ahlgren (2017). Combined Lagrange multiplier test for ARCH in vector autoregressive models. *Econometrics and Statistics 1*, 62–84.

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

Cenesizoglu, T. and A. Timmermann (2008). Is the distribution of stock returns predictable? *SSRN Working Paper*.

Choi, Y., S. Jacewitz, and J. Park (2016). A reexamination of stock return predictability. *Journal of Econometrics 192*, 168–189.

Cochrane, J. (1992). Explaining the variance of price-dividend ratios. *Review of Financial Studies 5*, 243–280.

Cochrane, J. (2008). The dog that did not bark: A defense of return predictabilty. *Review of Financial Studies 21*, 1533–1575.

Dufour, J.-M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics. *Journal of Econometrics 133*, 443–477.

Dufour, J.-M. and L. Khalaf (2001). Monte Carlo test methods in econometrics. In B. Baltagi (Ed.), *A Companion to Theoretical Econometrics*, pp. 494–510. Basil Blackwell, Oxford, UK.

Dufour, J.-M., L. Khalaf, and M. Voia (2015). Finite-sample resampling-based combined hypothesis tests, with applications to serial correlation and predictability. *Communications in Statistics – Simulation and Computation, 44*, 2329–2347.

Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics 28*, 181–187.

Dybvig, P. and H. Zhang (2018). That is not my dog: Why doesn't the log dividend-price ratio seem to predict future log returns or log dividend growth? *SSRN Working Paper*.

Ernst, M. (2004). Permutation methods: A basis for exact inference. *Statistical Science 19*, 676–685.

Fama, E. and K. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics 22*, 3–25.

Folks, J. (1984). Combination of independent tests. In P. Krishnaiah and P. Sen (Eds.), *Handbook of Statistics 4: Nonparametric Methods*, pp. 113–121. North-Holland, Amsterdam.

Goyal, A. and I. Welch (2003). Predicting the equity premium with dividend ratios. *Management Science 49*, 639–654.

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

Gupta, A. and D. Nagar (2000). *Matrix Variate Distributions*. Chapman & Hall/CRC, Boca Raton.

Hall, P. and S. Sheather (1988). On the distribution of a Studentized quantile. *Journal of the Royal Statistical Society. Series B 50*, 381–391.

Hansen, L. and R. Hodrick (1980). Forward exchanger rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy 88*, 829–853.

Hjalmarsson, E. (2011). New methods for inference in long-horizon regressions. *Journal of Financial and Quantitative Analysis 46*, 815–839.

Hsu, J. (1996). *Multiple Comparisons: Theory and Methods*. Chapman & Hall/CRC, Boca Raton.

Jondeau, E. and M. Rockinger (2006). Optimal portfolio allocation under higher moments. *European Financial Management 12*, 29–55.

Kim, T.-H. and H. White (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters 1*, 56–73.

Kiviet, J. (2011). Monte Carlo simulation for econometricians. *Foundations and Trends in Econometrics 5*, 1–181.

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

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–49.

Koenker, R. and J. Machado (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association 94*(448), 1296–1310.

Koenker, R. and Z. Xiao (2006). Quantile autoregression. *Journal of the American Statistical Association 101*, 980–990.

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

Lanne, M. (2002). Testing the predictability of stock returns. *Review of Economics and Statistics 84*, 407–415.

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

Lehmann, E. and J. Romano (2005). *Testing Statistical Hypotheses, Third Edition*. Springer, New York.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics 74*, 209–235.

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

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

Mankiw, N. and M. Shapiro (1986). Do we reject too often?: Small sample properties of tests of rational expectation models. *Economics Letters 20*, 139–145.

Maynard, A., K. Shimotsu, and Y. Wang (2011). Inference in predictive quantile regressions. *University of Guelph Working Paper*.

Nankervis, J. and N. Savin (1985). Testing the autoregressive parameter with the t statistic. *Journal of Econometrics 27*, 143–161.

Nankervis, J. and N. Savin (1988). The Student's t approximation in a stationary first order autoregressive model. *Econometrica 56*, 119–145.

Nankervis, J. and N. Savin (1996). The level and power of the bootstrap t test in the AR(1) model with trend. *Journal of Business and Economic Statistics 14*, 161–168.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Park, C. (2010). When does the dividend-price ratio predict stock returns? *Journal of Empirical Finance 17*, 81–101.

Phillips, P. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica 82*, 1177–1195.

Polk, C., S. Thompson, and T. Vuolteenaho (2006). Cross-sectional forecasts of the equity premium. *Journal of Financial Economics 81*, 101–141.

Randles, R. and D. Wolfe (1979). *Introduction to the Theory of Nonparametric Statistics*. Wiley, New York.

Rayner, R. (1990). Bootstrapping p values and power in the first-order autoregression: a Monte Carlo investigation. *Journal of Business and Economic Statistics 8*, 251–263.

Robertson, D. and S. Wright (2006). Dividends, total cash flow to shareholders, and predictive return regressions. *Review of Economics and Statistics 88*, 91–99.

Savin, N. (1984). Multiple hypothesis testing. In Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics*, pp. 827–879. North-Holland, Amsterdam.

Schwert, W. (1989). Why does stock market volatility change over time? *Journal of Finance 44*, 1115–1153.

Siddiqui, M. (1960). Distribution of quantiles in samples from a bivariate population. *Journal of Research of the National Bureau of Standards. Section B 64*, 145–150.

Stambaugh, R. (1999). Predictive regressions. *Journal of Financial Economics 54*, 375–421.

Tanaka, K. (1983). Asymptotic expansions associated with the AR(1) model with unknown mean. *Econometrica 51*, 1221–1231.

Taylor, J. (2017). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business and Economic Statistics*, DOI: 10.1080/07350015.2017.1281815.

Taylor, S. (1986). *Modelling Financial Time Series*. John Wiley and Sons, Chichester, UK.

Tippett, L. (1931). *The Methods of Statistics*. Williams and Norgate, London.

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

Valkanov, R. (2003). Long-horizon regressions: Theoretical results and applications. *Journal of Financial Economics 68*, 210–232.

Van Nieuwerburgh, S. and R. Koijen (2009). Financial economics, return predictability and market efficiency. In R. Meyers (Ed.), *Complex Systems in Finance and Econometrics*, pp. 353–360. Springer, New York.

Westfall, P. and S. Young (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. Wiley, New York.

Wilkinson, B. (1951). A statistical consideration in psychological research. *Psychology Bulletin 48*, 156–158.

Table 1: Standard t- and IVX-QR tests at individual quantiles: size

	$\gamma = 0$									$\gamma = -0.95$								
$\tau =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\phi = 0.5$																		
t-test	11.4	9.0	7.2	6.3	6.6	5.8	6.3	8.2	12.1	10.0	7.4	5.4	6.5	6.8	6.0	7.2	9.2	10.8
IVX-QR ($\delta = 0.1$)	6.7	7.8	9.6	9.9	9.7	9.2	9.4	7.8	7.4	12.4	7.2	7.1	5.1	4.8	5.7	6.5	8.7	10.3
IVX-QR ($\delta = 0.2$)	10.7	8.2	6.3	6.1	6.2	6.4	5.5	8.2	8.2	9.9	7.7	6.2	7.0	6.1	7.7	7.5	8.7	9.9
IVX-QR ($\delta = 0.3$)	11.8	7.5	5.7	7.2	6.4	6.3	6.8	9.2	8.8	9.5	7.7	7.2	7.3	6.1	6.6	8.1	8.1	9.4
IVX-QR ($\delta = 0.4$)	11.4	7.6	7.0	6.1	5.7	5.9	7.7	8.1	8.8	9.3	7.7	7.2	7.0	6.2	6.9	7.7	8.1	10.9
IVX-QR ($\delta = 0.5$)	12.1	8.5	7.0	5.5	5.5	5.9	7.2	8.1	10.2	9.9	7.4	6.9	7.0	6.1	6.7	8.6	7.6	10.4
IVX-QR ($\delta = 0.6$)	11.6	8.1	7.1	6.6	6.8	6.1	7.0	7.5	10.4	9.2	7.8	6.8	5.1	6.2	6.4	7.3	7.8	10.1
IVX-QR ($\delta = 0.7$)	11.0	8.5	7.4	5.8	5.7	5.6	6.8	7.7	11.0	10.2	7.8	6.7	4.9	6.5	6.6	7.6	8.7	10.8
IVX-QR ($\delta = 0.8$)	11.3	8.7	7.1	6.3	6.0	5.9	6.5	7.5	11.3	10.4	7.6	6.1	5.6	6.0	6.4	6.9	7.7	10.3
IVX-QR ($\delta = 0.9$)	11.1	8.2	6.9	6.5	7.2	6.3	6.8	8.3	11.9	10.4	7.7	5.6	6.1	6.1	6.6	7.0	8.0	10.8
$\phi = 0.95$																		
t-test	10.0	8.0	8.0	7.4	6.6	7.3	5.7	8.9	12.2	10.0	8.6	9.2	8.7	8.2	8.7	9.7	10.0	10.7
IVX-QR ($\delta = 0.1$)	8.0	9.9	10.8	9.7	11.6	11.4	12.4	10.2	7.6	12.1	7.0	6.9	5.6	5.6	5.7	6.8	8.7	11.1
IVX-QR ($\delta = 0.2$)	11.5	9.6	7.9	6.8	6.5	6.9	6.0	7.9	11.8	9.8	7.7	5.8	5.7	4.8	6.6	7.1	8.1	9.2
IVX-QR ($\delta = 0.3$)	12.3	8.6	7.5	6.5	5.7	7.4	6.1	7.6	10.7	9.9	7.1	7.0	5.9	5.6	6.5	7.1	8.4	9.3
IVX-QR ($\delta = 0.4$)	12.0	9.0	7.7	5.9	6.7	7.7	6.5	9.1	10.2	8.0	7.8	6.9	5.5	6.4	6.4	7.4	7.7	10.2
IVX-QR ($\delta = 0.5$)	11.6	7.1	6.0	6.2	7.4	7.8	7.8	9.1	8.9	7.0	7.9	7.2	5.1	6.0	6.6	7.7	6.9	10.4
IVX-QR ($\delta = 0.6$)	9.6	7.8	6.4	7.6	6.6	7.1	7.4	9.6	10.7	6.8	7.8	7.1	5.8	5.9	7.3	8.7	7.3	10.3
IVX-QR ($\delta = 0.7$)	9.3	9.0	5.7	6.9	7.6	6.6	6.3	9.8	9.6	8.4	7.3	6.7	6.6	6.1	6.4	7.7	8.3	9.4
IVX-QR ($\delta = 0.8$)	10.9	7.8	6.7	7.9	8.1	6.7	7.1	8.9	10.1	9.0	6.7	6.7	6.4	6.8	5.6	7.2	8.4	9.3
IVX-QR ($\delta = 0.9$)	10.5	8.1	7.2	7.9	8.3	6.6	7.3	9.4	10.4	9.8	6.7	6.6	6.3	6.9	6.1	7.0	8.9	9.3
$\phi = 0.99$																		
t-test	11.2	9.8	8.0	6.3	6.8	7.4	5.7	6.9	10.1	14.2	12.4	12.2	14.0	13.7	12.8	12.9	12.4	12.0
IVX-QR ($\delta = 0.1$)	7.6	10.9	11.9	12.2	13.3	13.0	11.8	10.9	7.8	9.9	10.5	7.9	7.5	6.5	5.1	6.7	7.5	11.1
IVX-QR ($\delta = 0.2$)	10.9	7.9	8.0	5.7	5.7	7.1	5.5	7.2	9.0	11.1	7.3	6.6	6.2	6.4	4.9	6.4	8.4	9.0
IVX-QR ($\delta = 0.3$)	10.7	7.4	7.1	6.9	4.8	5.2	5.1	7.2	10.3	8.9	8.0	7.6	6.0	7.7	7.6	7.7	8.5	10.2
IVX-QR ($\delta = 0.4$)	9.6	8.5	6.6	5.7	5.0	4.3	5.3	6.9	12.0	8.5	6.9	6.7	6.1	7.2	7.1	7.5	7.2	10.9
IVX-QR ($\delta = 0.5$)	10.3	7.4	5.9	6.2	4.6	4.0	4.9	7.4	10.0	8.3	7.8	6.4	6.1	6.0	5.5	7.0	6.3	10.2
IVX-QR ($\delta = 0.6$)	9.6	7.8	7.7	6.1	5.2	5.7	6.0	6.7	9.9	9.0	7.1	6.9	5.3	5.6	6.3	6.4	6.5	9.3
IVX-QR ($\delta = 0.7$)	10.0	7.2	7.1	5.7	6.7	6.7	6.3	8.7	11.5	8.8	9.5	6.1	6.1	5.9	5.9	6.2	6.3	9.1
IVX-QR ($\delta = 0.8$)	10.4	7.5	6.6	6.6	6.9	6.3	6.8	9.2	10.7	10.4	8.7	6.9	5.6	6.6	6.1	6.7	6.2	8.5
IVX-QR ($\delta = 0.9$)	10.3	7.8	7.3	7.2	6.5	6.6	6.7	8.0	11.0	11.9	8.4	7.2	5.8	6.4	6.6	6.8	7.3	9.6
$\phi = 1$																		
t-test	10.2	7.9	6.2	6.9	5.4	5.0	7.3	8.0	12.4	15.2	17.6	18.2	17.1	20.1	21.0	19.1	17.3	15.7
IVX-QR ($\delta = 0.1$)	7.7	8.5	9.4	11.8	11.0	10.6	9.5	8.3	8.2	10.7	9.4	6.9	6.9	5.8	7.4	7.5	8.8	11.9
IVX-QR ($\delta = 0.2$)	9.6	8.5	6.4	5.1	4.8	5.9	6.5	7.9	11.4	10.0	8.5	7.9	7.6	6.9	7.4	7.7	7.5	9.9
IVX-QR ($\delta = 0.3$)	10.8	7.5	5.9	5.0	5.1	4.9	7.3	8.1	11.2	8.4	7.0	6.8	7.3	6.6	6.9	8.1	7.4	10.1
IVX-QR ($\delta = 0.4$)	10.5	7.6	5.6	5.2	6.4	5.9	7.6	8.2	11.2	9.8	8.3	6.5	7.1	7.2	6.8	8.0	7.4	9.0
IVX-QR ($\delta = 0.5$)	10.9	7.6	6.6	5.1	5.7	5.0	6.7	8.4	9.9	10.2	8.7	7.0	8.4	7.4	6.8	7.1	7.7	12.9
IVX-QR ($\delta = 0.6$)	9.9	6.9	6.9	6.8	5.8	5.1	6.0	7.5	9.6	11.3	8.7	7.0	8.5	7.3	7.2	7.6	7.2	12.6
IVX-QR ($\delta = 0.7$)	9.6	8.2	5.9	6.8	5.6	6.2	6.3	7.7	9.8	10.7	8.8	8.8	9.0	8.0	7.7	8.2	8.0	12.6
IVX-QR ($\delta = 0.8$)	10.5	6.5	7.2	7.4	4.8	5.6	6.3	7.4	11.7	10.4	10.1	9.0	8.7	7.4	7.5	8.0	8.7	11.1
IVX-QR ($\delta = 0.9$)	11.2	6.9	6.9	7.0	5.3	6.7	7.2	7.2	11.3	10.7	10.7	9.6	9.9	9.4	8.9	9.3	11.4	11.1
$\phi = 1.01$																		
t-test	10.6	8.0	6.7	5.8	4.2	4.3	6.6	9.7	10.7	15.7	18.1	15.2	16.7	17.9	17.2	18.3	17.6	17.9
IVX-QR ($\delta = 0.1$)	7.6	8.6	11.4	12.5	11.7	11.9	9.9	9.4	5.8	10.5	7.8	7.3	6.2	5.2	5.8	6.6	7.9	12.2
IVX-QR ($\delta = 0.2$)	9.1	8.6	6.8	6.2	6.1	6.6	6.5	9.0	11.0	9.8	7.8	6.7	5.1	5.6	6.3	6.3	7.3	9.7
IVX-QR ($\delta = 0.3$)	10.0	8.2	6.8	7.7	6.3	7.0	7.5	8.9	9.6	10.2	8.4	8.1	6.5	6.7	6.2	7.0	7.8	9.3
IVX-QR ($\delta = 0.4$)	10.4	8.7	8.1	7.7	7.1	7.8	7.2	9.0	11.7	9.1	8.3	7.6	7.3	6.6	6.9	8.6	8.7	9.9
IVX-QR ($\delta = 0.5$)	11.5	8.2	8.8	7.9	7.1	6.4	7.7	10.0	12.3	9.7	8.6	8.8	7.8	8.8	8.2	9.0	9.2	10.0
IVX-QR ($\delta = 0.6$)	11.1	9.4	7.6	7.6	5.9	7.2	6.9	9.5	11.5	11.5	8.7	9.2	9.6	10.8	11.1	9.8	10.3	12.2
IVX-QR ($\delta = 0.7$)	11.4	8.8	8.3	7.3	6.3	6.4	7.8	8.2	11.2	11.6	9.9	9.6	10.7	11.6	12.1	11.7	12.0	12.8
IVX-QR ($\delta = 0.8$)	11.3	8.3	8.0	6.8	6.4	6.1	7.1	8.2	12.4	12.8	10.8	11.6	12.2	12.6	12.5	14.9	13.6	12.9
IVX-QR ($\delta = 0.9$)	10.3	8.4	7.1	6.7	5.5	6.3	6.6	8.3	10.6	14.9	13.3	12.6	13.9	14.3	15.0	14.9	14.7	14.4

Notes: This table reports the empirical size (in percentage) of the standard t-test and the IVX-QR test of Lee (2016). The nominal level is $\alpha = 5\%$.

Table 2: LMC test across quantiles and horizons: size

	$\gamma = 0$										$\gamma = -0.95$									
$\tau =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	all τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	all τ
$\phi = 0.5$																				
LMC ($h = 1$)	5.2	5.3	4.8	4.1	5.3	4.8	5.0	5.0	5.7	5.5	5.5	4.1	4.0	4.5	4.8	4.5	5.9	5.7	5.1	5.1
LMC ($h = 2$)	4.5	5.2	5.6	6.4	5.2	5.1	4.0	5.1	4.3	4.4	4.6	5.7	6.0	5.1	5.7	5.3	5.2	5.9	5.1	6.3
LMC ($h = 3$)	4.3	4.6	6.1	5.4	7.0	5.5	4.4	6.2	5.2	6.6	5.3	4.9	5.9	6.4	6.3	6.4	6.9	5.7	5.3	5.5
LMC (all h)	5.6	5.7	5.3	4.9	5.6	5.1	5.2	5.7	5.0	5.0	4.4	4.3	5.8	5.9	5.7	6.3	6.4	5.2	4.7	4.8
$\phi = 0.95$																				
LMC ($h = 1$)	5.0	5.2	5.9	4.8	5.2	5.2	3.5	5.5	6.0	6.2	5.5	5.5	6.3	5.5	5.3	7.1	6.2	5.3	4.9	4.8
LMC ($h = 2$)	4.7	6.5	5.1	5.8	5.1	5.1	5.7	5.0	6.5	6.1	5.7	6.2	5.3	5.1	6.6	5.7	6.6	4.7	6.2	5.8
LMC ($h = 3$)	5.6	5.2	5.1	5.7	4.3	4.2	5.0	6.7	5.7	4.8	6.1	6.6	5.7	6.0	6.4	6.3	6.6	5.7	5.0	5.9
LMC (all h)	4.7	5.6	4.9	6.1	5.0	5.2	5.2	6.6	6.4	5.4	6.2	6.1	6.2	5.8	5.3	5.7	6.1	5.6	5.4	5.5
$\phi = 0.99$																				
LMC ($h = 1$)	4.7	6.5	6.1	4.5	5.5	5.2	4.2	4.8	5.1	4.9	7.2	7.0	6.6	9.2	9.8	7.9	8.0	7.0	6.4	7.2
LMC ($h = 2$)	5.5	5.1	5.2	5.8	6.1	5.2	4.5	3.7	5.4	5.8	8.6	7.2	9.5	8.2	10.1	8.6	6.6	8.0	6.2	7.7
LMC ($h = 3$)	6.0	5.6	5.2	4.9	4.5	5.2	3.7	4.4	5.3	5.4	6.3	7.6	8.5	7.5	8.1	10.4	8.3	8.1	6.5	7.8
LMC (all h)	6.4	6.4	5.7	5.6	5.0	6.0	4.3	4.1	5.1	4.6	6.6	7.5	8.4	8.3	8.5	10.3	8.4	8.6	7.1	7.2
$\phi = 1$																				
LMC ($h = 1$)	5.8	5.4	5.1	4.9	5.2	5.0	4.7	5.5	5.2	5.3	7.5	11.6	11.3	12.9	14.6	11.9	11.2	10.5	8.2	11.4
LMC ($h = 2$)	5.0	4.1	5.2	5.9	4.7	4.7	3.9	4.3	5.0	4.1	8.9	11.5	11.7	15.8	14.5	12.9	11.9	10.1	10.4	12.8
LMC ($h = 3$)	5.3	5.2	3.8	4.6	5.6	5.2	4.4	4.9	4.3	5.0	8.7	11.4	13.7	12.7	14.6	13.9	13.1	10.7	11.0	12.1
LMC (all h)	5.2	4.6	4.1	4.8	5.9	4.8	4.5	4.4	4.6	5.1	8.5	11.2	13.4	14.0	15.1	13.9	12.3	11.2	10.3	11.7
$\phi = 1.01$																				
LMC ($h = 1$)	7.1	5.0	4.2	4.9	4.8	4.2	4.9	5.1	4.3	5.7	7.9	9.3	12.1	9.9	11.6	11.7	10.4	9.9	8.9	11.2
LMC ($h = 2$)	5.8	5.2	4.8	4.1	5.7	5.4	5.4	4.8	4.7	5.6	9.4	11.3	11.1	12.8	13.5	12.7	12.1	10.9	8.2	13.6
LMC ($h = 3$)	3.5	5.7	4.6	5.0	4.5	5.8	5.5	4.8	5.0	4.6	8.6	10.5	13.1	12.4	14.0	13.1	11.9	10.2	9.6	11.6
LMC (all h)	5.2	5.5	4.7	5.1	4.5	6.5	5.3	5.4	5.5	5.0	8.4	11.5	13.3	13.4	14.1	13.0	11.7	11.2	9.7	12.0

Notes: This table reports the empirical size (in percentage) of the LMC test. The nominal level is $\alpha = 5\%$.

Table 3: MMC test across quantiles and horizons: size

	$\gamma = 0$										$\gamma = -0.95$									
$\tau =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	all τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	all τ
$\phi = 0.5$																				
MMC ($h = 1$)	1.0	0.9	1.0	0.9	1.0	0.9	0.7	1.0	1.3	1.4	1.3	0.5	0.9	1.1	1.0	1.7	1.0	0.8	1.1	0.8
MMC ($h = 2$)	0.7	1.5	1.5	1.5	1.1	1.3	0.8	0.9	0.4	0.9	0.5	0.9	1.0	1.4	1.3	1.4	1.9	1.8	1.0	0.9
MMC ($h = 3$)	1.0	0.9	0.5	1.4	1.2	0.9	1.2	1.0	0.6	0.5	0.9	0.8	1.1	1.3	1.5	1.3	1.0	0.7	0.8	1.0
MMC (all h)	0.4	0.9	0.9	1.4	1.0	1.0	1.2	1.4	1.2	0.9	0.7	1.0	1.4	1.3	1.7	1.4	1.5	1.0	0.7	0.9
$\phi = 0.95$																				
MMC ($h = 1$)	1.2	1.5	1.4	1.8	2.1	2.1	0.7	2.0	1.8	1.6	1.0	1.0	0.7	0.7	1.3	0.9	0.9	1.2	0.8	0.8
MMC ($h = 2$)	1.0	1.3	1.1	1.3	1.2	1.6	1.4	2.1	1.9	1.5	1.2	0.9	1.3	0.9	0.5	0.8	1.1	0.4	1.1	0.5
MMC ($h = 3$)	0.7	1.8	1.6	1.2	1.4	1.2	1.4	1.7	1.3	1.9	0.9	1.1	0.7	0.8	0.8	0.8	0.6	0.9	1.1	1.0
MMC (all h)	0.9	1.8	1.7	1.4	1.5	1.2	1.8	1.7	1.7	2.0	0.7	1.0	1.1	0.7	0.7	0.9	1.0	0.7	1.2	0.7
$\phi = 0.99$																				
MMC ($h = 1$)	1.8	2.1	1.7	1.2	1.2	1.5	1.5	1.4	1.9	1.9	1.6	1.0	1.3	1.2	1.4	1.7	0.9	1.3	1.5	1.4
MMC ($h = 2$)	2.0	1.9	1.5	1.7	1.6	2.1	1.1	1.1	0.8	1.2	1.0	0.7	0.9	1.3	1.2	1.7	1.2	2.1	1.6	0.9
MMC ($h = 3$)	2.1	1.3	1.0	1.8	1.3	1.3	1.2	1.1	1.7	1.2	1.0	1.3	1.1	1.1	1.3	1.3	1.0	1.6	1.7	1.4
MMC (all h)	1.4	1.9	1.5	2.0	1.3	1.5	1.1	1.2	1.3	1.3	0.7	0.9	0.9	1.1	1.6	1.4	1.1	1.5	1.7	1.2
$\phi = 1$																				
MMC ($h = 1$)	0.4	0.9	1.1	1.3	0.8	1.1	1.1	0.9	1.2	0.8	0.8	2.1	2.1	1.1	2.1	1.2	1.1	1.5	1.3	1.0
MMC ($h = 2$)	0.4	0.7	1.3	0.7	1.1	1.2	0.7	0.5	0.4	0.7	1.3	1.5	1.4	1.3	1.4	1.3	1.5	1.3	1.3	1.2
MMC ($h = 3$)	0.8	0.7	0.7	0.8	1.2	1.5	1.0	1.4	1.1	1.0	1.0	1.1	1.6	1.4	1.6	1.4	1.2	1.0	1.2	1.1
MMC (all h)	0.7	0.8	0.9	1.0	1.0	1.2	0.8	0.7	1.1	1.0	0.9	1.1	1.4	1.4	1.5	1.3	1.3	1.1	0.9	1.2
$\phi = 1.01$																				
MMC ($h = 1$)	2.4	1.1	1.2	1.7	1.4	1.4	1.5	1.3	1.4	1.8	1.5	1.3	1.6	2.0	1.4	1.1	2.0	2.0	2.2	2.0
MMC ($h = 2$)	1.2	1.5	1.6	1.0	1.6	2.3	1.4	1.4	1.3	1.6	1.9	1.6	1.2	2.1	2.0	1.6	2.2	2.3	1.3	1.7
MMC ($h = 3$)	1.0	1.4	1.0	1.2	1.1	1.7	1.6	1.6	1.3	1.0	1.8	1.8	1.6	1.1	1.4	2.3	1.7	1.6	1.6	1.4
MMC (all h)	1.3	1.8	1.0	1.2	1.4	1.9	1.3	1.8	1.4	1.1	1.8	1.9	1.5	1.7	1.7	2.4	2.0	1.6	1.3	1.3

Notes: This table reports the empirical size (in percentage) of the MMC test. The nominal level is $\alpha = 5\%$.

Table 4: Standard t- and IVX-QR tests at individual quantiles: size-adjusted power

	$\gamma=0$									$\gamma=-0.95$								
$\tau=$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\phi=0.5$																		
t-test	7.5	22.0	42.3	64.8	71.4	61.6	43.6	22.7	8.2	11.8	32.8	51.6	66.7	76.1	66.3	49.9	27.8	12.9
IVX-QR ($\delta=0.1$)	2.5	4.7	8.9	17.8	25.2	18.6	8.7	4.1	2.0	1.2	4.0	12.3	25.5	34.2	24.3	10.7	3.9	1.8
IVX-QR ($\delta=0.2$)	4.7	6.9	6.4	8.7	4.3	8.1	7.8	5.7	6.9	5.6	6.3	7.0	5.9	4.1	6.8	7.9	6.3	6.4
IVX-QR ($\delta=0.3$)	5.4	11.9	16.4	13.0	10.7	14.5	14.0	9.7	8.2	7.9	14.6	15.0	13.2	11.8	13.5	14.4	11.7	7.8
IVX-QR ($\delta=0.4$)	6.7	13.6	22.1	25.4	23.2	24.9	19.6	14.8	9.6	10.5	21.7	22.9	23.9	26.1	27.0	24.4	18.1	9.2
IVX-QR ($\delta=0.5$)	6.2	17.9	28.7	38.2	36.3	37.6	26.4	17.0	9.2	13.4	22.4	30.1	39.5	39.0	36.9	30.6	22.5	10.8
IVX-QR ($\delta=0.6$)	7.8	18.3	31.8	47.1	45.9	46.1	34.8	18.7	8.0	12.8	25.2	34.8	50.4	48.9	45.5	34.3	23.4	12.3
IVX-QR ($\delta=0.7$)	7.9	19.7	35.7	54.4	57.2	53.1	37.2	19.9	8.7	13.3	28.2	40.4	57.9	56.5	54.7	38.1	24.4	10.7
IVX-QR ($\delta=0.8$)	7.8	21.7	37.9	57.5	62.1	55.5	40.1	21.8	8.6	12.3	31.3	43.9	57.6	62.3	59.0	39.4	24.6	10.7
IVX-QR ($\delta=0.9$)	7.9	21.3	39.1	57.5	64.2	58.3	40.7	21.7	10.0	11.7	31.0	46.7	59.8	65.6	62.0	41.7	24.0	13.3
$\phi=0.95$																		
t-test	30.7	45.0	54.2	62.6	67.7	63.0	57.0	43.6	27.6	33.0	50.8	58.1	66.7	75.1	66.9	57.3	51.7	39.9
IVX-QR ($\delta=0.1$)	2.1	5.2	9.8	19.2	25.5	19.3	9.8	5.5	3.8	2.1	5.1	11.4	23.7	32.9	26.2	12.2	5.6	1.7
IVX-QR ($\delta=0.2$)	6.6	4.9	6.7	6.1	4.5	6.2	7.3	5.3	4.6	4.7	5.9	7.0	5.6	6.0	4.7	5.5	5.2	5.9
IVX-QR ($\delta=0.3$)	6.3	8.3	8.4	6.7	5.1	7.3	10.3	7.3	7.3	6.6	7.9	8.5	7.5	5.1	6.7	7.3	8.7	8.8
IVX-QR ($\delta=0.4$)	7.6	12.6	9.6	9.6	6.8	8.4	11.6	9.2	6.7	12.6	9.9	10.0	11.0	6.8	9.4	10.6	14.6	9.9
IVX-QR ($\delta=0.5$)	11.1	17.8	16.1	10.4	8.4	12.7	13.1	12.6	12.4	18.3	14.8	16.7	15.9	9.3	12.0	13.4	18.6	14.2
IVX-QR ($\delta=0.6$)	16.5	23.3	22.5	15.2	14.0	17.9	20.0	17.9	12.9	25.3	22.1	22.4	22.7	13.8	18.5	20.3	25.7	21.2
IVX-QR ($\delta=0.7$)	23.3	28.2	30.3	23.5	17.5	26.0	28.4	23.2	19.9	29.8	32.5	31.6	29.0	20.1	24.3	28.3	31.5	25.8
IVX-QR ($\delta=0.8$)	24.2	37.1	38.6	34.2	25.0	32.9	36.5	31.6	22.9	31.6	41.0	39.7	39.2	32.7	38.4	39.4	38.7	31.8
IVX-QR ($\delta=0.9$)	25.5	41.2	43.3	44.9	39.6	44.6	43.8	37.4	24.7	33.4	46.8	49.9	48.7	43.0	46.1	48.2	44.1	34.9
$\phi=0.99$																		
t-test	60.5	70.4	72.2	65.5	56.9	67.3	73.0	71.6	61.3	61.5	71.5	72.1	65.6	61.0	64.9	68.1	67.5	61.6
IVX-QR ($\delta=0.1$)	5.1	8.3	11.4	21.8	29.2	21.0	11.3	6.7	4.0	2.3	4.7	12.6	25.0	33.6	26.5	13.2	7.0	2.3
IVX-QR ($\delta=0.2$)	4.7	5.1	5.5	6.6	4.4	4.5	6.5	4.7	5.4	4.8	3.8	4.5	6.2	4.4	6.8	5.5	4.2	4.9
IVX-QR ($\delta=0.3$)	6.7	6.7	6.4	6.0	6.4	6.9	8.9	7.0	5.6	7.1	7.0	5.9	7.9	4.2	5.6	5.6	6.4	6.2
IVX-QR ($\delta=0.4$)	8.3	9.5	10.2	9.7	6.7	9.6	9.1	9.6	6.9	11.5	12.7	11.5	10.4	5.8	8.6	8.7	9.5	7.0
IVX-QR ($\delta=0.5$)	11.0	13.9	12.8	11.2	8.6	11.2	13.6	11.7	9.4	14.2	15.2	14.6	12.1	7.6	12.3	10.6	16.3	12.7
IVX-QR ($\delta=0.6$)	15.5	18.9	17.7	16.5	10.0	13.1	19.1	21.6	13.7	21.9	23.2	20.7	18.4	10.6	14.7	18.3	21.0	19.1
IVX-QR ($\delta=0.7$)	24.2	28.0	25.8	24.9	14.7	19.2	26.2	25.9	24.8	28.1	30.2	29.7	22.7	16.3	20.8	25.8	31.9	28.1
IVX-QR ($\delta=0.8$)	35.4	40.4	37.4	29.6	21.0	28.2	35.2	37.1	32.9	34.3	39.2	39.7	33.9	21.9	28.0	36.7	44.7	37.6
IVX-QR ($\delta=0.9$)	43.7	48.5	46.6	37.3	30.1	36.5	46.9	50.5	41.9	45.8	53.0	48.9	44.3	29.9	38.7	45.0	48.5	44.5
$\phi=1$																		
t-test	75.8	81.0	81.7	66.0	51.0	65.6	82.4	82.2	77.6	77.1	76.4	71.7	61.8	52.2	61.1	73.2	77.4	77.2
IVX-QR ($\delta=0.1$)	5.4	8.6	16.1	28.9	36.6	30.2	15.6	10.0	5.1	3.3	10.9	20.2	35.9	45.7	34.9	20.7	9.8	4.9
IVX-QR ($\delta=0.2$)	6.5	4.3	5.7	6.1	5.8	5.8	4.5	5.7	4.6	4.8	5.7	6.4	7.2	5.0	5.1	7.2	6.2	6.3
IVX-QR ($\delta=0.3$)	7.5	5.6	6.3	6.9	5.1	6.5	6.1	6.6	6.1	6.4	7.1	5.5	5.7	5.6	7.1	8.1	6.4	4.7
IVX-QR ($\delta=0.4$)	6.4	6.6	7.4	8.0	5.4	8.3	10.1	8.1	7.4	9.3	8.1	7.7	7.7	6.3	6.9	10.2	9.0	6.1
IVX-QR ($\delta=0.5$)	8.1	11.3	10.4	9.7	7.9	11.6	14.5	13.0	10.1	12.3	12.7	9.3	10.7	7.5	11.1	12.7	12.7	10.6
IVX-QR ($\delta=0.6$)	13.3	14.2	15.8	12.8	9.7	12.9	17.9	20.0	12.8	19.0	18.0	16.9	15.3	13.0	15.1	18.1	17.1	19.0
IVX-QR ($\delta=0.7$)	20.4	22.0	21.7	17.4	14.3	19.3	25.5	27.4	21.5	27.0	23.2	24.7	22.1	16.9	21.8	26.1	26.0	26.3
IVX-QR ($\delta=0.8$)	28.2	30.6	31.1	24.8	20.1	25.7	35.6	34.6	29.6	36.4	33.3	32.8	27.5	23.6	29.5	37.2	37.3	35.0
IVX-QR ($\delta=0.9$)	38.7	44.4	39.0	35.8	25.4	38.1	47.5	44.9	37.8	46.0	47.3	42.4	37.7	32.3	39.4	47.4	47.4	45.2
$\phi=1.01$																		
t-test	90.2	94.2	93.9	80.4	61.8	82.3	93.1	94.1	92.5	92.8	92.7	88.5	76.8	66.3	76.6	89.3	92.6	92.1
IVX-QR ($\delta=0.1$)	11.7	24.1	38.1	53.9	64.2	56.4	38.6	24.0	10.6	9.7	29.0	44.9	65.5	72.7	67.0	44.4	24.6	9.2
IVX-QR ($\delta=0.2$)	7.0	7.7	7.8	7.7	6.0	6.0	7.7	9.7	6.7	6.1	7.4	7.8	7.4	5.7	7.0	6.1	7.1	7.1
IVX-QR ($\delta=0.3$)	15.6	17.6	18.9	13.8	7.3	11.8	16.9	15.7	18.3	17.5	20.0	15.1	9.9	6.5	10.3	13.3	16.2	13.4
IVX-QR ($\delta=0.4$)	26.3	29.1	29.9	16.7	10.5	16.2	28.2	30.3	27.1	27.2	30.5	25.5	16.3	9.7	15.5	24.5	27.0	27.3
IVX-QR ($\delta=0.5$)	38.1	39.6	39.4	24.4	16.8	26.8	39.8	41.8	37.3	40.5	43.8	35.5	22.5	15.3	23.8	33.1	41.2	44.7
IVX-QR ($\delta=0.6$)	49.4	51.5	52.1	37.1	20.2	36.0	48.5	55.4	52.1	54.2	54.4	47.1	33.2	20.1	33.9	45.2	55.0	53.3
IVX-QR ($\delta=0.7$)	54.0	64.0	60.6	46.5	30.0	47.6	62.6	64.2	61.7	65.0	64.9	56.8	43.3	26.0	46.3	60.3	65.8	66.4
IVX-QR ($\delta=0.8$)	67.4	74.4	71.0	56.0	39.0	56.6	71.3	73.2	72.9	72.4	73.5	65.9	53.6	37.7	56.5	69.5	72.4	73.7
IVX-QR ($\delta=0.9$)	76.5	82.3	79.6	63.6	47.4	65.6	78.3	80.2	78.6	77.5	80.5	74.2	61.9	49.5	64.0	76.5	79.3	78.7

Notes: This table reports the size-adjusted power (in percentage) of the standard t-test, and the IVX-QR test for different values of δ. The nominal level is $\alpha=5\%$.

Table 5: LMC test across quantiles and horizons: size-adjusted power

	$\gamma = 0$										$\gamma = -0.95$									
$\tau =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	all τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	all τ
$\phi = 0.5$																				
LMC ($h = 1$)	6.4	20.7	40.6	63.0	71.9	60.8	36.5	20.5	7.1	59.9	11.3	29.9	49.7	65.8	78.0	66.7	46.5	26.5	12.3	68.4
LMC ($h = 2$)	7.9	16.3	23.6	38.0	43.4	38.6	28.1	15.6	7.8	31.5	14.7	38.7	68.8	86.9	92.3	87.7	69.1	34.6	11.5	84.2
LMC ($h = 3$)	8.4	14.5	16.0	24.7	26.3	24.3	20.6	10.4	6.0	15.9	14.3	39.2	60.2	75.1	80.9	74.0	56.9	34.5	13.5	69.8
LMC (all h)	6.3	17.6	33.8	56.9	67.3	54.7	30.7	16.0	6.5	48.0	16.7	39.6	69.2	86.6	93.8	86.3	64.5	36.9	14.6	83.9
$\phi = 0.95$																				
LMC ($h = 1$)	26.7	42.9	50.1	62.4	67.7	62.9	54.6	42.2	23.8	75.6	34.5	47.1	52.4	64.0	74.6	64.8	52.0	47.7	37.8	85.8
LMC ($h = 2$)	24.7	34.8	44.7	51.9	55.9	51.2	44.6	38.0	19.3	60.5	31.4	47.6	61.8	76.6	83.4	78.7	63.4	53.0	29.8	90.0
LMC ($h = 3$)	18.8	31.6	38.7	42.5	48.5	47.7	38.5	29.6	18.0	52.5	26.3	47.6	65.8	80.8	84.2	79.8	66.3	49.3	32.1	90.4
LMC (all h)	24.5	38.8	53.1	56.7	66.1	59.7	47.9	35.1	22.2	62.7	29.3	52.6	65.8	84.5	89.1	84.0	69.5	54.0	35.0	91.0
$\phi = 0.99$																				
LMC ($h = 1$)	59.8	66.8	70.7	65.5	56.9	65.4	69.2	68.5	58.9	88.5	59.4	68.0	68.6	63.2	62.0	62.4	62.6	65.2	59.3	91.2
LMC ($h = 2$)	51.3	62.0	63.3	56.6	48.5	54.9	61.1	62.7	50.9	79.7	49.1	60.9	60.8	64.2	69.9	64.0	60.9	55.7	53.7	93.8
LMC ($h = 3$)	44.9	55.1	56.6	50.5	47.2	50.5	57.5	57.0	44.8	72.1	50.3	52.9	59.5	66.9	70.8	67.2	58.2	51.7	44.9	93.4
LMC (all h)	50.6	64.2	67.6	63.1	59.1	64.0	68.2	64.4	52.0	82.0	56.5	64.9	66.5	71.3	73.5	70.8	68.0	59.0	51.3	94.4
$\phi = 1$																				
LMC ($h = 1$)	75.0	80.9	77.9	66.5	52.1	63.4	80.4	80.7	75.7	93.2	75.9	76.2	66.4	60.1	54.9	59.4	67.9	76.7	73.6	96.0
LMC ($h = 2$)	68.0	78.0	69.7	56.3	47.9	55.0	72.7	75.9	67.5	90.1	63.1	65.1	59.4	59.8	60.4	59.8	62.2	66.9	64.7	94.6
LMC ($h = 3$)	60.6	67.9	65.8	50.4	41.3	47.2	63.4	68.8	64.0	81.6	55.0	59.9	56.6	59.1	62.5	58.2	56.8	59.4	56.9	95.9
LMC (all h)	68.2	77.1	76.1	63.9	51.7	61.2	75.4	78.4	70.8	87.6	67.4	65.3	65.2	62.2	64.6	63.9	69.2	70.7	68.1	97.1
$\phi = 1.01$																				
LMC ($h = 1$)	90.5	93.5	93.4	80.7	62.8	82.0	92.9	93.1	92.2	97.6	92.9	91.7	86.0	76.4	67.0	75.9	87.9	91.8	91.4	99.6
LMC ($h = 2$)	86.9	89.5	86.6	71.7	59.0	73.4	87.0	90.6	88.8	95.4	85.2	84.8	78.0	70.4	68.0	71.3	80.0	86.8	87.8	99.7
LMC ($h = 3$)	86.5	84.5	80.4	62.9	55.9	66.2	81.7	86.7	84.0	94.2	80.2	79.4	74.0	67.2	68.4	68.6	75.5	80.1	83.0	99.7
LMC (all h)	88.0	91.0	91.5	79.5	68.9	80.1	89.8	90.4	88.7	95.1	87.9	87.3	85.5	78.2	75.5	77.5	86.2	87.3	87.4	99.8

Notes: This table reports the size-adjusted power (in percentage) of the LMC test. The nominal level is $\alpha = 5\%$.

Table 6: MMC test across quantiles and horizons: power

	$\gamma = 0$										$\gamma = -0.95$									
$\tau =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	all τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	all τ
$\phi = 0.5$																				
MMC ($h = 1$)	1.3	8.4	22.3	48.1	62.3	46.5	20.0	8.4	2.1	43.7	3.7	11.2	29.4	52.1	66.9	52.8	28.6	12.1	3.3	51.4
MMC ($h = 2$)	1.2	5.8	12.9	22.0	24.9	21.8	12.2	5.4	1.9	10.9	3.5	20.7	50.5	78.0	86.6	76.7	51.5	20.1	2.4	70.9
MMC ($h = 3$)	1.3	2.8	8.3	12.4	15.0	11.3	8.0	3.6	1.9	5.1	4.5	17.0	37.5	59.5	66.4	59.8	37.0	15.0	4.2	46.6
MMC (all h)	1.0	6.3	17.8	40.2	52.6	38.9	16.9	5.3	1.4	29.7	4.4	20.2	48.3	76.4	85.8	75.5	49.2	18.8	3.1	65.6
$\phi = 0.95$																				
MMC ($h = 1$)	13.8	29.1	38.3	51.2	58.4	49.2	37.2	27.8	13.6	59.9	17.3	29.8	38.7	52.5	62.2	52.2	38.9	30.4	19.6	68.5
MMC ($h = 2$)	9.4	25.2	34.4	41.1	43.3	38.4	32.7	21.3	9.2	38.8	15.4	31.9	46.9	63.8	75.6	66.4	50.4	32.0	16.2	77.1
MMC ($h = 3$)	7.4	18.8	26.2	30.4	35.0	31.2	25.7	17.5	7.7	25.7	13.0	33.5	50.8	68.6	74.9	69.3	54.0	34.1	15.8	75.4
MMC (all h)	9.6	24.9	35.1	46.1	51.9	46.2	34.1	21.6	10.2	38.7	13.9	35.2	53.5	71.3	79.0	72.5	56.6	34.6	17.2	76.0
$\phi = 0.99$																				
MMC ($h = 1$)	44.9	58.8	61.0	53.9	48.8	55.2	59.0	57.9	45.1	78.1	49.5	57.3	57.8	55.5	54.0	54.4	55.5	54.9	48.3	86.1
MMC ($h = 2$)	37.1	49.2	52.3	45.3	40.3	44.8	50.9	47.6	34.9	65.2	39.6	50.0	53.8	57.8	63.4	57.3	52.1	48.4	38.0	89.8
MMC ($h = 3$)	28.9	39.6	43.4	37.2	34.8	38.5	43.2	41.4	26.8	52.9	35.8	46.6	51.6	60.4	63.9	59.9	50.8	43.7	35.0	87.8
MMC (all h)	35.8	50.2	55.9	51.3	46.6	52.5	53.9	49.4	35.3	62.4	42.9	53.7	58.6	63.7	66.8	62.7	57.7	51.9	40.3	89.1
$\phi = 1$																				
MMC ($h = 1$)	59.0	69.3	65.9	51.4	37.0	50.0	66.4	70.5	59.5	82.4	63.6	67.8	60.5	50.0	44.1	53.0	61.3	68.4	64.3	92.3
MMC ($h = 2$)	48.2	58.9	54.8	41.4	31.1	38.2	54.1	59.0	49.9	70.4	51.7	55.5	54.0	51.7	50.5	52.4	53.5	57.2	54.2	91.9
MMC ($h = 3$)	38.9	50.0	45.9	33.8	27.4	33.8	44.1	50.0	40.6	60.6	44.8	50.7	52.0	50.4	51.7	51.7	50.5	49.3	44.0	88.8
MMC (all h)	47.3	61.9	56.0	45.4	35.7	44.2	57.4	60.6	48.6	68.7	52.7	58.9	59.2	54.9	54.8	57.3	57.7	59.5	56.1	92.0
$\phi = 1.01$																				
MMC ($h = 1$)	84.8	88.6	88.1	77.1	51.1	72.7	87.0	88.6	85.2	93.6	87.8	88.5	83.6	70.8	59.7	72.8	85.5	89.5	87.7	99.0
MMC ($h = 2$)	79.0	83.1	78.9	61.3	47.5	64.7	80.7	84.3	81.6	89.6	81.0	83.2	76.4	66.5	61.1	67.9	77.6	82.9	82.2	99.4
MMC ($h = 3$)	72.2	76.6	71.6	52.8	44.8	56.0	73.9	80.4	74.9	85.5	75.7	77.9	70.7	63.2	63.5	65.2	72.9	77.7	78.1	98.8
MMC (all h)	79.4	85.4	82.3	68.4	54.8	69.9	83.3	86.7	81.3	88.7	83.2	84.9	82.6	74.3	70.1	74.8	83.8	85.6	83.6	99.0

Notes: This table reports the power (in percentage) of the MMC test. No size adjustments were made because the probability of a Type I error with the MMC tests is $\leq \alpha$. The nominal level is $\alpha = 5\%$.

Table 7: LMC and MMC tests: size under departures from the maintained assumptions

$\tau =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	all τ
AR(2)										
LMC ($h = 1$)	6.1	5.9	7.6	7.1	7.2	7.9	7.0	5.4	5.7	6.3
LMC ($h = 2$)	6.7	6.3	7.2	7.5	9.0	8.0	6.9	6.5	4.3	6.5
LMC ($h = 3$)	7.8	6.7	7.1	7.9	8.2	8.1	7.0	7.8	5.6	7.1
LMC (all h)	7.6	7.4	6.5	8.6	7.6	7.4	6.9	7.9	4.6	6.7
MMC ($h = 1$)	1.0	0.6	0.5	0.8	0.7	0.7	1.0	0.7	0.7	0.3
MMC ($h = 2$)	1.0	1.3	0.7	0.6	0.7	0.3	0.6	0.5	0.4	0.5
MMC ($h = 3$)	0.7	0.8	0.8	1.0	0.5	0.9	1.1	0.4	0.9	1.0
MMC (all h)	1.1	0.6	0.8	1.0	0.5	1.0	1.0	0.3	0.6	1.0
GARCH(1,1)										
LMC ($h = 1$)	15.0	12.7	9.7	7.7	9.5	8.9	11.5	14.2	15.1	17.1
LMC ($h = 2$)	13.1	10.3	8.4	8.3	8.6	8.3	11.5	10.2	11.9	13.7
LMC ($h = 3$)	11.0	11.0	10.1	8.2	7.9	10.0	10.2	10.9	11.3	13.3
LMC (all h)	11.4	11.6	9.3	8.1	8.7	9.7	10.4	11.3	11.7	14.4
MMC ($h = 1$)	4.0	2.3	1.2	0.9	1.0	0.8	2.0	2.9	3.3	3.1
MMC ($h = 2$)	1.9	1.6	1.3	0.6	1.1	0.6	1.2	2.0	2.3	2.2
MMC ($h = 3$)	1.6	0.7	0.9	1.3	1.1	0.8	1.0	1.2	2.4	1.6
MMC (all h)	1.4	1.3	1.2	1.3	1.2	0.9	1.1	1.8	3.0	2.1
AR(2) + GARCH(1,1)										
LMC ($h = 1$)	14.8	12.4	11.4	8.3	7.5	8.2	10.0	12.4	14.6	18.1
LMC ($h = 2$)	12.7	10.3	8.0	9.1	7.7	7.8	10.0	11.9	10.7	13.4
LMC ($h = 3$)	12.0	10.5	8.3	7.7	7.7	8.9	9.7	9.6	9.3	11.7
LMC (all h)	13.0	10.8	8.9	8.2	7.7	9.1	9.6	10.1	11.5	12.7
MMC ($h = 1$)	4.9	3.7	2.2	1.4	0.7	1.1	1.8	3.1	4.7	5.4
MMC ($h = 2$)	2.9	2.5	1.3	1.2	0.7	1.2	1.3	2.6	2.1	2.8
MMC ($h = 3$)	2.4	1.4	0.9	0.7	0.6	1.8	1.1	1.5	2.0	2.3
MMC (all h)	2.4	1.9	0.9	1.0	0.5	1.4	1.2	2.6	2.5	2.6

Notes: This table reports the empirical size (in percentage) of the LMC and MMC tests under violations of the maintained assumptions, *i.e.*, AR(1) predictor model and exchangeable errors. The nominal level is $\alpha = 5\%$.

Table 8: Distribution-free tests for median predictability: comparisons

	$\gamma = 0$					$\gamma = -0.95$				
$\phi =$	0.50	0.95	0.99	1.00	1.01	0.50	0.95	0.99	1.00	1.01
Size ($\beta_1 = 0$)										
$S(\hat{\beta}_0)$	3.3	4.9	4.2	3.3	1.4	7.5	6.7	6.4	5.2	2.7
$W(\hat{\beta}_0)$	4.1	4.4	5.0	4.1	1.5	5.4	5.8	5.9	5.3	2.8
S	0.4	0.4	0.4	0.2	0.1	0.5	0.5	0.2	0.1	0.2
W	1.2	1.5	1.3	0.3	0.1	2.1	1.9	0.8	0.6	0.8
LMC	4.3	4.8	3.9	5.5	4.6	5.4	5.6	9.0	10.9	11.6
MMC	1.3	1.4	0.8	1.7	1.8	1.0	1.2	1.4	1.8	3.9
Power ($\beta_1 = 0.05$)										
$S(\hat{\beta}_0)$	8.8	28.1	44.7	44.5	46.6	8.6	23.5	49.3	50.5	59.7
$W(\hat{\beta}_0)$	8.4	30.3	46.8	49.5	68.3	9.6	36.4	66.1	66.4	78.5
S	2.2	5.8	12.0	11.9	8.1	2.0	4.9	10.8	13.6	11.1
W	5.3	13.2	21.5	18.5	11.4	6.3	13.4	30.3	29.5	21.3
LMC	9.5	33.9	61.4	77.2	92.5	9.1	46.7	82.5	87.5	97.5
MMC	2.8	18.4	44.2	61.7	87.8	2.7	12.1	57.1	82.2	96.1
Power ($\beta_1 = 0.1$)										
$S(\hat{\beta}_0)$	21.9	68.9	83.6	75.1	66.9	16.9	65.4	93.2	82.2	71.3
$W(\hat{\beta}_0)$	22.7	71.5	85.6	79.5	87.9	25.9	90.0	98.1	88.0	93.5
S	4.0	31.1	49.1	39.4	20.9	6.4	36.6	61.8	50.8	29.5
W	12.1	47.5	61.7	47.6	27.4	18.1	69.0	82.9	66.4	37.2
LMC	23.9	83.8	97.0	97.2	99.6	27.7	95.7	100.0	99.9	100.0
MMC	7.0	64.0	90.0	93.0	98.5	9.1	70.3	98.3	99.7	100.0

Notes: This table compares the empirical size and power (in percentage) of the Campbell and Dufour (1997) tests for median predictability with the proposed LMC and MMC tests. The power calculations for the $S(\hat{\beta}_0)$, $W(\hat{\beta}_0)$, and LMC tests are based on size adjustments. The nominal level is $\alpha = 5\%$.

Table 9: Summary statistics and AR(1) estimates

	r	*d/p*	*e/p*	*btm*	*dfy*	*tms*	*tbl*
Panel A: Monthly data							
mean	5.646	-3.589	-2.830	0.504	0.010	0.018	0.048
median	9.492	-3.520	-2.867	0.446	0.009	0.019	0.049
stdev.	14.945	0.402	0.436	0.263	0.005	0.015	0.032
skew.	-0.405	-0.227	-0.749	0.720	1.682	-0.335	0.621
kurtosis	4.733	2.252	5.995	2.604	6.977	2.729	3.860
autocorr.	0.042	0.994	0.989	0.994	0.968	0.957	0.990
AR(1) estimates							
$\hat{\phi}$		0.995	0.998	1.003	0.982	0.967	1.002
$\hat{\sigma}(\hat{\phi})$		(0.004)	(0.005)	(0.003)	(0.007)	(0.007)	(0.001)
99% CI_ϕ		[0.985, 1.007]	[0.976, 1.016]	[0.992, 1.013]	[0.961, 1.005]	[0.938, 1.001]	[0.998, 1.005]
Panel B: Weekly data							
mean	2.592	-3.589	-2.819	0.505	0.010	0.016	0.048
median	9.352	-3.518	-2.864	0.453	0.009	0.016	0.049
stdev	15.669	0.401	0.390	0.262	0.005	0.013	0.032
skew.	-0.324	-0.243	0.096	0.722	1.746	-0.328	0.631
kurtosis	7.726	2.263	3.040	2.626	7.471	2.705	3.902
autocorr.	-0.043	0.999	0.998	0.999	0.993	0.991	0.998
AR(1) estimates							
$\hat{\phi}$		0.998	0.998	0.999	1.000	0.996	1.000
$\hat{\sigma}(\hat{\phi})$		(0.001)	(0.001)	(0.001)	(0.008)	(0.002)	(0.000)
99% CI_ϕ		[0.995, 1.001]	[0.995, 1.002]	[0.996, 1.002]	[0.976, 1.014]	[0.992, 1.000]	[0.999, 1.001]

Notes: This table presents some summary statistics for the excess returns on the S&P value-weighted index and six predictor variables. The data covers the period from January 1962 to December 2015. The mean/median and the standard deviation for percentage monthly (weekly) excess returns are annualized by multiplying by 12 (52) and $\sqrt{12}$ ($\sqrt{52}$), respectively. The predictor AR(1) persistence parameter estimate $\hat{\phi}$ is obtained by LAD and the numbers in parentheses are the associated standard errors. The numbers in brackets show the 99% confidence intervals for ϕ, obtained according to (18).

Table 10: Monthly return predictability: LMC tests

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	all τ
Predictor = d/p												
$h = 1$	0.91	0.68	0.97	0.61	0.61	1.00	0.49	0.49	0.34	0.82	0.25	0.90
$h = 3$	0.50	0.82	0.90	0.95	0.82	0.89	0.16	0.29	0.06	0.14	0.33	0.35
$h = 12$	0.42	0.42	0.32	0.76	0.78	0.84	0.81	0.71	0.39	0.38	0.54	0.71
$h = 60$	0.79	0.75	0.55	0.37	0.43	0.56	0.47	0.66	0.84	0.99	0.93	0.84
$h = 120$	0.41	0.66	0.80	0.76	0.73	0.78	0.65	0.67	0.59	0.54	0.21	0.56
all h	0.50	0.75	0.82	0.67	0.66	0.82	0.71	0.70	0.65	0.59	0.26	0.64
Predictor = e/p												
$h = 1$	0.57	0.12	0.77	0.68	0.44	0.78	0.93	0.98	0.78	0.30	0.73	0.58
$h = 3$	0.14	0.43	0.92	0.52	0.56	0.76	0.72	0.36	0.62	0.79	0.70	0.54
$h = 12$	0.73	0.27	0.24	0.90	0.84	0.78	1.00	0.78	0.49	0.84	0.77	0.65
$h = 60$	0.33	0.46	**0.03**	0.16	0.63	0.92	0.99	0.82	0.65	0.76	0.93	0.23
$h = 120$	0.09	0.15	0.29	0.19	0.56	0.62	0.67	0.35	0.57	0.66	0.26	0.23
all h	0.11	0.17	0.20	0.21	0.65	0.75	0.83	0.42	0.67	0.80	0.35	0.26
Predictor $= btm$												
$h = 1$	0.97	0.68	0.85	0.50	0.13	0.29	0.85	0.68	0.67	0.99	**0.02**	**0.03**
$h = 3$	0.66	0.92	0.58	0.48	0.27	0.83	0.95	0.67	0.44	0.21	0.69	0.67
$h = 12$	0.85	0.85	0.84	0.94	0.99	0.96	1.00	0.99	0.86	0.65	0.86	0.98
$h = 60$	0.94	0.83	0.69	0.84	0.93	0.97	1.00	0.95	0.89	0.63	0.93	0.93
$h = 120$	0.40	0.46	0.60	0.76	0.92	0.94	0.88	0.86	0.86	0.99	0.92	0.71
Predictor $= dfy$												
$h = 1$	0.43	0.10	0.20	0.76	0.68	0.58	**0.04**	0.08	**0.01**	**0.01**	**0.01**	**0.01**
$h = 3$	0.39	0.33	0.27	0.37	0.81	**0.05**	**0.03**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**
$h = 12$	0.55	0.91	0.80	0.36	0.10	**0.01**	**0.03**	0.07	0.28	0.09	0.31	0.14
$h = 60$	**0.02**	**0.01**	0.09	0.13	0.11	0.13	0.22	0.12	0.28	0.78	0.92	0.09
$h = 120$	0.29	**0.02**	0.17	0.17	0.48	0.55	0.35	0.10	0.07	0.50	0.81	0.17
all h	0.09	**0.02**	0.20	0.24	0.22	0.24	0.43	0.16	0.10	0.62	0.48	0.21
Predictor $= tms$												
$h = 1$	0.57	0.10	**0.02**	0.21	0.21	0.17	0.21	0.45	0.26	0.26	0.10	0.20
$h = 3$	0.97	0.24	**0.05**	**0.05**	0.10	0.10	0.11	0.26	**0.02**	0.09	0.23	0.10
$h = 12$	0.18	0.21	**0.02**	0.08	0.07	0.23	0.23	0.24	0.25	0.26	0.49	0.09
$h = 60$	0.52	0.43	0.15	0.20	0.06	**0.01**	**0.02**	**0.03**	**0.04**	0.41	0.57	0.06
$h = 120$	0.52	0.11	0.07	0.08	0.08	0.10	0.27	0.49	0.52	0.83	0.85	0.41
all h	0.72	0.15	0.11	0.17	0.12	**0.01**	**0.03**	**0.05**	0.17	0.70	0.86	0.14
Predictor $= tbl$												
$h = 1$	0.95	0.23	**0.01**	0.06	**0.02**	**0.02**	0.23	1.00	0.88	0.96	0.99	0.08
$h = 3$	0.92	0.61	**0.04**	**0.02**	**0.02**	**0.03**	0.07	0.85	0.91	0.97	0.70	0.06
$h = 12$	0.77	0.30	**0.03**	0.14	0.21	0.35	0.43	0.54	0.75	0.90	0.70	0.15
$h = 60$	0.63	0.59	0.36	0.57	0.63	0.86	0.88	0.96	0.83	0.73	0.94	0.82
$h = 120$	0.21	0.32	0.49	0.63	0.75	0.59	0.58	0.51	0.33	0.06	**0.04**	0.12
all h	0.22	0.39	0.56	0.75	0.78	0.71	0.70	0.58	0.39	0.09	**0.04**	0.14

Notes: This table shows the individual and joint p-values over prediction horizons $h = 1, 3, 12, 60, 120$ months and quantile levels τ ranging from 0.05 to 0.95. The entries in bold are instances of statistical significance at the 5% level.

Table 11: Monthly return predictability: MMC tests

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	all τ
Predictor $= d/p$												
$h = 1$	0.99	0.96	1.00	0.96	0.97	1.00	0.91	0.88	0.75	0.95	0.49	1.00
$h = 3$	0.81	0.98	1.00	1.00	1.00	1.00	0.66	0.84	0.34	0.42	0.51	0.92
$h = 12$	0.88	0.88	0.87	1.00	1.00	1.00	1.00	1.00	0.94	0.73	0.84	1.00
$h = 60$	0.97	0.97	0.94	0.90	0.96	0.99	0.94	1.00	1.00	0.99	0.98	1.00
$h = 120$	0.83	0.93	1.00	1.00	1.00	0.98	0.98	0.97	0.95	0.85	0.58	0.94
all h	0.90	0.96	1.00	1.00	0.98	1.00	1.00	0.98	0.97	0.89	0.62	0.96
Predictor $= e/p$												
$h = 1$	0.66	0.26	0.90	0.86	0.65	0.88	0.99	1.00	0.97	0.52	0.84	0.83
$h = 3$	0.22	0.67	0.97	0.75	0.82	0.90	0.92	0.60	0.75	0.89	0.83	0.74
$h = 12$	0.87	0.48	0.44	0.98	0.95	0.96	1.00	0.94	0.77	0.93	0.87	0.82
$h = 60$	0.55	0.63	0.13	0.49	0.81	0.96	1.00	0.96	0.83	0.89	0.97	0.49
$h = 120$	0.23	0.28	0.47	0.36	0.74	0.82	0.86	0.71	0.77	0.81	0.45	0.46
all h	0.25	0.32	0.38	0.45	0.82	0.88	0.92	0.78	0.84	0.92	0.55	0.47
Predictor $= btm$												
$h = 1$	1.00	0.75	0.95	0.58	0.12	0.30	0.86	0.76	0.66	1.00	**0.03**	**0.03**
$h = 3$	0.73	0.95	0.57	0.50	0.28	0.85	0.99	0.69	0.43	0.21	0.70	0.70
$h = 12$	0.88	0.87	0.89	0.96	1.00	0.99	1.00	1.00	0.94	0.75	0.91	1.00
$h = 60$	0.94	0.85	0.65	0.83	0.94	0.99	1.00	0.98	0.94	0.77	0.97	0.96
$h = 120$	0.49	0.45	0.67	0.82	0.94	0.96	0.94	0.90	0.95	1.00	0.95	0.83
all h	0.57	0.51	0.72	0.84	0.96	1.00	0.99	0.97	0.97	0.95	0.95	0.88
Predictor $= dfy$												
$h = 1$	0.58	0.18	0.32	0.83	0.75	0.70	0.08	0.12	**0.01**	**0.02**	**0.03**	**0.03**
$h = 3$	0.51	0.53	0.46	0.62	0.86	0.14	0.06	**0.01**	**0.02**	**0.03**	**0.02**	**0.04**
$h = 12$	0.70	0.99	0.88	0.56	0.27	0.10	0.15	0.18	0.35	0.23	0.46	0.35
$h = 60$	0.14	0.06	0.21	0.23	0.22	0.25	0.48	0.29	0.56	0.85	0.98	0.24
$h = 120$	0.60	0.24	0.42	0.43	0.68	0.73	0.61	0.33	0.30	0.74	0.92	0.55
all h	0.38	0.26	0.46	0.45	0.44	0.45	0.66	0.37	0.34	0.82	0.73	0.59
Predictor $= tms$												
$h = 1$	0.61	0.14	0.06	0.26	0.31	0.25	0.28	0.53	0.36	0.48	0.12	0.25
$h = 3$	0.98	0.32	0.08	0.11	0.14	0.13	0.15	0.34	0.05	0.17	0.35	0.21
$h = 12$	0.28	0.30	0.05	0.15	0.14	0.30	0.28	0.40	0.44	0.40	0.62	0.15
$h = 60$	0.64	0.53	0.29	0.31	0.17	**0.04**	0.08	0.09	0.22	0.64	0.77	0.19
$h = 120$	0.68	0.31	0.35	0.32	0.31	0.34	0.46	0.67	0.75	0.95	0.92	0.68
all h	0.84	0.38	0.39	0.37	0.30	0.09	0.15	0.22	0.38	0.85	0.96	0.47
Predictor $= tbl$												
$h = 1$	1.00	0.37	**0.04**	0.09	**0.04**	**0.04**	0.26	1.00	0.92	0.98	0.98	0.19
$h = 3$	0.98	0.76	0.10	0.06	**0.03**	0.05	0.21	0.90	0.96	1.00	0.81	0.10
$h = 12$	0.85	0.46	0.08	0.23	0.28	0.41	0.59	0.66	0.90	0.93	0.84	0.27
$h = 60$	0.68	0.65	0.45	0.59	0.61	0.91	0.94	1.00	0.93	0.89	0.99	0.88
$h = 120$	0.33	0.48	0.58	0.70	0.81	0.68	0.71	0.60	0.42	0.22	0.12	0.28
all h	0.37	0.55	0.67	0.79	0.79	0.81	0.86	0.68	0.49	0.24	0.13	0.32

Notes: See Table 10.

Table 12: Weekly return predictability: LMC tests

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	all τ
Predictor = d/p												
$h = 1$	0.71	0.92	0.89	0.87	0.99	0.33	0.98	0.77	0.89	0.87	0.69	0.97
$h = 4$	0.87	0.48	0.91	0.91	0.62	0.38	0.74	0.88	0.92	0.47	0.13	0.42
$h = 12$	0.92	0.87	0.89	0.74	0.55	0.55	1.00	0.48	0.36	0.25	0.34	0.51
$h = 24$	0.71	0.94	0.81	0.98	0.75	0.69	0.90	0.88	0.27	0.22	0.25	0.57
$h = 52$	0.77	0.53	0.52	0.97	0.86	0.81	0.88	0.97	0.82	0.62	0.74	0.86
all h	0.88	0.60	0.54	0.99	0.85	0.82	0.96	0.80	0.52	0.44	0.46	0.81
Predictor = e/p												
$h = 1$	0.10	0.40	0.31	0.33	0.20	0.13	0.24	0.33	0.71	0.65	0.19	0.41
$h = 4$	0.44	0.11	0.53	0.51	0.61	0.98	0.53	0.37	0.48	0.56	0.52	0.34
$h = 12$	0.52	0.41	0.83	0.90	0.91	0.98	0.39	0.23	0.19	0.27	0.75	0.47
$h = 24$	0.23	0.56	0.98	0.93	0.84	0.87	0.90	0.74	0.29	0.43	0.62	0.75
$h = 52$	0.85	0.59	0.39	0.91	0.96	1.00	0.97	0.90	0.92	0.92	0.78	0.87
all h	0.40	0.67	0.43	0.99	0.94	0.96	0.79	0.65	0.61	0.58	0.85	0.89
Predictor $= btm$												
$h = 1$	0.69	0.18	0.17	**0.05**	0.29	0.94	0.40	0.95	0.94	0.79	0.65	0.09
$h = 4$	0.55	0.57	0.29	0.17	0.14	0.06	0.40	0.71	0.89	0.39	0.19	0.18
$h = 12$	0.93	0.50	0.20	0.11	0.10	0.17	0.46	0.98	0.56	0.35	0.48	0.24
$h = 24$	0.87	0.63	0.33	0.37	0.33	0.23	0.34	0.63	0.69	0.49	0.43	0.59
$h = 52$	0.91	0.90	0.73	0.53	0.57	0.43	0.42	0.65	0.81	0.99	0.91	0.84
all h	0.97	0.80	0.45	0.40	0.43	0.44	0.44	0.68	0.87	0.70	0.67	0.85
Predictor $= dfy$												
$h = 1$	**0.03**	**0.01**	**0.01**	**0.02**	0.22	0.39	**0.04**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**
$h = 4$	0.48	0.11	0.14	0.27	0.96	0.45	**0.05**	**0.01**	**0.01**	**0.01**	**0.01**	**0.01**
$h = 12$	0.43	0.16	0.36	0.59	0.98	0.51	0.11	**0.02**	**0.01**	**0.01**	**0.01**	**0.01**
$h = 24$	0.87	0.62	0.94	0.74	0.59	0.29	0.19	0.10	**0.01**	0.06	**0.04**	0.06
$h = 52$	0.37	0.67	0.94	0.78	0.38	0.31	0.09	**0.03**	0.09	0.13	0.34	0.14
all h	0.40	0.29	0.70	0.71	0.40	0.31	0.09	**0.03**	0.06	0.14	0.08	0.14
Predictor $= tms$												
$h = 1$	0.07	**0.01**	0.32	**0.05**	0.21	0.50	0.50	0.34	0.09	0.09	0.41	0.09
$h = 4$	**0.03**	**0.03**	**0.04**	0.08	0.26	0.21	0.08	0.08	0.07	0.13	0.08	0.06
$h = 12$	0.43	**0.05**	**0.05**	**0.05**	0.11	0.13	0.25	0.13	0.11	0.08	0.35	0.09
$h = 24$	0.45	0.08	**0.03**	0.08	0.11	0.19	0.28	0.26	0.28	0.34	0.55	0.09
$h = 52$	0.26	0.06	**0.03**	**0.05**	0.10	0.24	0.33	0.40	0.54	0.45	0.14	0.10
all h	0.30	0.06	**0.03**	**0.05**	0.10	0.24	0.34	0.40	0.52	0.48	0.14	0.10
Predictor $= tbl$												
$h = 1$	0.85	**0.04**	**0.01**	**0.01**	**0.01**	0.15	0.42	0.49	0.77	0.60	0.34	**0.01**
$h = 4$	0.33	0.06	**0.02**	**0.01**	**0.01**	**0.01**	**0.01**	0.11	0.16	0.92	0.85	**0.01**
$h = 12$	0.39	0.11	**0.03**	**0.01**	**0.01**	**0.01**	0.09	0.23	0.56	0.79	0.94	**0.04**
$h = 24$	0.74	0.13	**0.05**	**0.01**	**0.03**	**0.02**	0.07	0.28	0.89	0.89	0.87	**0.05**
$h = 52$	0.47	**0.04**	**0.02**	0.06	0.07	0.14	0.28	0.54	0.75	0.76	0.82	0.11
all h	0.58	**0.05**	**0.02**	**0.05**	0.08	0.11	0.26	0.47	0.80	0.90	0.99	0.12

Notes: This table shows the individual and joint p-values over prediction horizons $h = 1, 4, 12, 24, 52$ weeks and quantile levels τ ranging from 0.05 to 0.95. The entries in bold are instances of statistical significance at the 5% level.

Table 13: Weekly return predictability: MMC tests

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	all τ
Predictor = d/p												
$h = 1$	0.86	0.98	0.99	1.00	1.00	0.88	1.00	0.95	0.98	0.98	0.83	1.00
$h = 4$	0.99	0.93	1.00	1.00	1.00	0.97	1.00	1.00	1.00	0.83	0.28	1.00
$h = 12$	0.98	0.99	1.00	1.00	1.00	1.00	1.00	0.99	0.90	0.55	0.61	1.00
$h = 24$	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89	0.63	0.55	1.00
$h = 52$	0.95	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.92	1.00
all h	0.99	0.97	0.99	1.00	1.00	1.00	1.00	1.00	0.99	0.88	0.81	1.00
Predictor = e/p												
$h = 1$	0.29	0.74	0.82	0.83	0.70	0.54	0.79	0.71	0.95	0.84	0.31	0.93
$h = 4$	0.77	0.43	0.97	0.99	0.99	1.00	0.96	0.91	0.89	0.86	0.75	0.94
$h = 12$	0.78	0.86	1.00	1.00	1.00	1.00	0.96	0.80	0.67	0.55	0.93	0.98
$h = 24$	0.60	0.92	1.00	1.00	1.00	1.00	1.00	1.00	0.84	0.77	0.89	1.00
$h = 52$	0.95	0.94	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	1.00
all h	0.80	0.97	0.94	1.00	1.00	1.00	1.00	1.00	0.96	0.92	0.98	1.00
Predictor = btm												
$h = 1$	0.90	0.44	0.50	0.08	0.56	1.00	0.69	1.00	0.99	0.89	0.77	0.21
$h = 4$	0.70	0.77	0.68	0.47	0.39	0.23	0.78	0.91	0.98	0.58	0.33	0.53
$h = 12$	0.96	0.71	0.45	0.27	0.30	0.51	0.81	1.00	0.78	0.52	0.65	0.55
$h = 24$	0.94	0.79	0.62	0.75	0.59	0.56	0.70	0.91	0.88	0.71	0.63	0.87
$h = 52$	0.96	0.95	0.92	0.84	0.85	0.75	0.81	0.91	0.94	1.00	1.00	0.99
all h	1.00	0.94	0.80	0.74	0.75	0.77	0.82	0.93	0.97	0.90	0.82	0.99
Predictor = dfy												
$h = 1$	0.19	**0.02**	**0.03**	0.07	0.30	0.50	0.17	**0.03**	**0.02**	**0.02**	**0.04**	0.06
$h = 4$	0.62	0.30	0.30	0.46	0.98	0.53	0.16	**0.04**	0.05	0.05	0.07	0.19
$h = 12$	0.66	0.44	0.54	0.73	1.00	0.61	0.23	0.10	0.08	0.13	0.10	0.37
$h = 24$	0.97	0.78	0.99	0.90	0.73	0.53	0.43	0.26	0.14	0.30	0.30	0.61
$h = 52$	0.70	0.86	0.98	0.86	0.57	0.47	0.34	0.28	0.37	0.41	0.59	0.78
all h	0.83	0.69	0.87	0.92	0.64	0.54	0.39	0.33	0.38	0.53	0.47	0.89
Predictor = tms												
$h = 1$	0.13	**0.04**	0.37	0.09	0.26	0.58	0.52	0.51	0.18	0.09	0.45	0.12
$h = 4$	0.05	**0.04**	0.09	0.10	0.30	0.20	0.10	0.11	0.11	0.14	0.10	0.12
$h = 12$	0.61	0.10	0.09	0.06	0.12	0.15	0.25	0.12	0.10	0.09	0.47	0.14
$h = 24$	0.55	0.24	0.05	0.07	0.14	0.20	0.20	0.22	0.23	0.41	0.54	0.13
$h = 52$	0.31	0.14	0.05	0.09	0.15	0.24	0.28	0.37	0.45	0.45	0.22	0.16
all h	0.33	0.15	0.05	0.09	0.15	0.25	0.28	0.38	0.47	0.47	0.24	0.16
Predictor = tbl												
$h = 1$	0.88	0.07	**0.01**	**0.01**	**0.02**	0.24	0.45	0.57	0.82	0.65	0.41	**0.02**
$h = 4$	0.35	0.12	0.07	**0.02**	**0.04**	0.05	0.06	0.16	0.25	0.99	0.91	0.06
$h = 12$	0.55	0.24	**0.04**	**0.04**	**0.03**	**0.04**	0.10	0.24	0.57	0.82	0.97	0.10
$h = 24$	0.80	0.16	0.09	**0.03**	0.05	0.06	0.12	0.28	0.89	0.90	0.81	0.13
$h = 52$	0.64	0.11	0.10	0.13	0.15	0.20	0.34	0.63	0.81	0.81	0.91	0.36
all h	0.76	0.11	0.10	0.12	0.15	0.18	0.32	0.54	0.89	0.91	0.99	0.37

Notes: See Table 12.

Table 14: Economic significance: Monthly data, $h = 1$

$\tau =$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
Excess return											
$\hat{Q}_\tau(r_t)$	-6.97%	-4.92%	-2.61%	-1.39%	-0.27%	0.80%	1.63%	2.72%	3.78%	5.37%	7.03%
Predictor = d/p											
$\hat{\beta}_1(\tau)$	0.002	0.005	0.001	0.003	-0.004	0.000	0.005	0.006	0.005	0.003	0.017
$R^2(\tau)$	0.006%	0.058%	0.022%	0.032%	0.052%	0.000%	0.058%	0.204%	0.202%	0.027%	0.914%
$\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]$	0.082%	0.197%	0.040%	0.107%	0.170%	0.003%	0.192%	0.234%	0.200%	0.127%	0.665%
$\frac{\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]}{\|\hat{Q}_\tau(r_t)\|}$	1.171%	4.011%	1.524%	7.694%	62.779%	0.345%	11.781%	8.610%	5.279%	2.368%	9.470%
Predictor = e/p											
$\hat{\beta}_1(\tau)$	0.010	0.013	0.002	0.002	-0.005	-0.001	0.000	0.000	0.001	-0.006	-0.002
$R^2(\tau)$	0.709%	0.645%	0.057%	0.031%	0.108%	0.008%	0.000%	0.001%	0.010%	0.253%	0.143%
$\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]$	0.433%	0.566%	0.107%	0.089%	0.226%	0.063%	0.019%	0.018%	0.039%	0.259%	0.084%
$\frac{\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]}{\|\hat{Q}_\tau(r_t)\|}$	6.213%	11.514%	4.104%	6.396%	83.571%	7.867%	1.144%	0.674%	1.031%	4.827%	1.201%
Predictor = btm											
$\hat{\beta}_1(\tau)$	-0.001	-0.012	-0.003	-0.008	-0.020	-0.011	-0.004	0.007	0.005	-0.001	**0.032**
$R^2(\tau)$	0.002%	0.111%	0.005%	0.135%	0.542%	0.202%	0.022%	0.109%	0.127%	0.002%	0.867%
$\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]$	0.020%	0.311%	0.076%	0.216%	0.522%	0.281%	0.118%	0.188%	0.142%	0.019%	0.833%
$\frac{\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]}{\|\hat{Q}_\tau(r_t)\|}$	0.291%	6.333%	2.931%	15.523%	193.043%	35.054%	7.258%	6.909%	3.764%	0.352%	11.857%
Predictor = dfy											
$\hat{\beta}_1(\tau)$	-1.268	-1.375	-0.793	-0.140	-0.268	0.197	0.903	1.184	**1.630**	**2.323**	**2.610**
$R^2(\tau)$	0.946%	1.145%	0.179%	0.011%	0.022%	0.060%	0.311%	0.813%	1.779%	3.579%	4.544%
$\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]$	0.579%	0.628%	0.362%	0.064%	0.122%	0.090%	0.412%	0.540%	0.744%	1.060%	1.191%
$\frac{\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]}{\|\hat{Q}_\tau(r_t)\|}$	8.300%	12.764%	13.868%	4.583%	45.138%	11.199%	25.308%	19.855%	19.665%	19.736%	16.954%
Predictor = tms											
$\hat{\beta}_1(\tau)$	0.312	0.486	0.385	0.193	0.226	0.158	0.161	0.158	0.123	0.228	0.443
$R^2(\tau)$	0.338%	0.858%	0.870%	0.278%	0.475%	0.314%	0.306%	0.101%	0.140%	0.415%	0.498%
$\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]$	0.466%	0.726%	0.575%	0.289%	0.337%	0.237%	0.240%	0.235%	0.184%	0.341%	0.661%
$\frac{\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]}{\|\hat{Q}_\tau(r_t)\|}$	6.680%	14.761%	22.033%	20.766%	124.671%	29.490%	14.745%	8.651%	4.866%	6.346%	9.411%
Predictor = tbl											
$\hat{\beta}_1(\tau)$	0.012	-0.133	**-0.177**	-0.157	**-0.174**	**-0.154**	-0.089	-0.003	0.009	-0.012	-0.003
$R^2(\tau)$	0.003%	0.313%	0.929%	0.427%	0.950%	0.488%	0.252%	0.002%	0.004%	0.016%	0.000%
$\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]$	0.038%	0.421%	0.559%	0.495%	0.549%	0.486%	0.281%	0.009%	0.029%	0.037%	0.008%
$\frac{\hat{\sigma}[\hat{Q}_\tau(r_t\|x_{t-1})]}{\|\hat{Q}_\tau(r_t)\|}$	0.541%	8.569%	21.425%	35.620%	203.026%	60.613%	17.229%	0.314%	0.768%	0.695%	0.117%

Notes: The top portion of this table reports the estimated unconditional quantiles $\hat{Q}_\tau(r_t)$ of r_t, the one-month excess returns on the S&P value-weighted index. Then for each predictor in turn, the table reports the estimated slope coefficient $\hat{\beta}_1(\tau)$ from the predictive quantile regression and the $R^2(\tau)$ goodness-of-fit measure developed by Koenker and Machado (1999) for quantile regressions. The reported ratios have in the numerator $\hat{\sigma}[\hat{Q}_\tau(r_t|x_{t-1})]$, which is the estimated standard deviation of the conditional quantile function, given the predictor; and the denominator is the absolute value of the corresponding sample quantile. The entries in bold are instances where the MMC predictability test (in Table 11) reveals $\hat{\beta}_1(\tau)$ to be significantly different from zero.