A Robust Inference for Predictive Expectile Regression: An IVX-Based Approach

Zongwu Cai
Department of Economics, University of Kansas
and
Wei Long
Department of Economics, Tulane University

March 27, 2026

Abstract

This paper develops a persistence-robust inferential framework for predictive expectile regression with highly persistent regressors. We combine expectile score equations with IVX instruments to construct an IVX-expectile estimator that preserves the distributional interpretation of expectile regression while regularizing the nonstandard effects of near-unit-root regressors, endogeneity, and conditional heteroscedasticity. For fixed expectile levels, we establish consistency and asymptotic normality of the estimator and show that the associated Wald statistic converges to a standard chi-square distribution. Simulation evidence indicates that the proposed procedure delivers accurate size for regressors with differential persistence, with only a modest local-power cost relative to conventional methods. In an application to monthly and quarterly U.S. stock return predictability, the method detects substantially asymmetric predictive ability across expectiles, showing that IVX-expectile regression provides a useful tool for studying heterogeneous predictive effects and downside tail risk when predictors are highly persistent.

Keywords: IVX inference; Persistent predictors; Predictive expectile regression; Stock return

1 Introduction

Forecasting asset returns with persistent state variables is a central problem in financial econometrics. The predictors that appear most economically interesting, such as valuation ratios, interest rates, and other macro-financial variables, are often also the most persistent, so the question of return predictability is inseparable from the problem of persistence-robust inference. Early studies document predictive relations using these state variables (Campbell and Shiller, 1988; Pesaran and Timmermann, 1995). Subsequent work shows why conventional inference can be misleading in this scenario: Stambaugh (1999) highlights the bias created by persistent and endogenous predictors, Campbell and Yogo (2006) develop tests tailored to nearly integrated regressors, and Welch and Goyal (2008) show that many apparently successful forecasting relations are unstable in practice. Together, these papers make clear that any credible study of stock return predictability must address predictors' persistence directly.

Most of this literature, however, studies predictability through the conditional mean. Quantile regression provides a natural way to allow predictive effects to vary across the conditional distribution, and it has become the standard distributional alternative to mean regression (Koenker and Bassett, 1978; Koenker, 2005). Expectile regression, introduced by Newey and Powell (1987) and interpreted by Efron (1991) through asymmetric squared loss, offers a different distributional perspective. Like quantiles, expectiles capture heterogeneous effects across the distribution; unlike quantiles, they remain sensitive to the magnitude of tail realizations and are based on a smooth objective function. This distinction is especially important in finance, where researchers care not only about whether predictors shift the tails, but also about how severely they do so.

Expectiles are also attractive because they have a special status in risk measurement; see, for example, the survey by Tian et al. (2019) for more discussions. Ziegel (2016) emphasizes that expectiles occupy a unique position among coherent risk measures because they are elicitable, while Bellini et al. (2014) links them to generalized quantiles and coherent tail-risk evaluation. This matters because Gneiting (2011) shows that elicitability is central for meaningful forecast evaluation, whereas expected shortfall (ES) on its own lacks this property although it is coherent risk measure. The applied literature has already exploited these features: Taylor (2008), Kuan et al. (2009), Cai et al. (2018), and Cai et al. (2025) use expectiles to model Value-at-Risk and ES dynamically, and Daouia et al. (2018) and Davison et al. (2023) show that expectiles are effective tools for extreme and heavy-tailed risk analysis. Accordingly, expectiles are not merely a smooth substitute for quantiles. Instead, they provide a distributional object that is economically meaningful about both tail location and tail severity.

Despite these advantages, predictive expectile regression with persistent covariates remains largely undeveloped in the literature. By contrast, predictive quantile regression has already addressed persistence and endogeneity through a sequence of contributions covering unit root dynamics, IVX-based estimation, heteroscedasticity, and unified inference (Koenker and Xiao, 2004, 2006; Lee, 2016; Fan and Lee, 2019; Cai et al., 2023; Maynard et al., 2024). In mean regression, the IVX literature shows how mildly integrated instruments can preserve low-frequency signal while restoring feasible inference with nearly integrated predictors (Magdalinos and Phillips, 2009; Kostakis et al., 2015; Phillips and Lee, 2016). So far, to the best of our knowledge, no analogous persistence-robust framework has been developed for predictive expectile regression, even though expectiles are especially well suited to asymmetric tail-risk questions for applied researchers.

This paper fills that gap by proposing an IVX-based framework for predictive expectile regression with highly persistent predictors. The method replaces the original persistent regressors in the expectile score equations with IVX-filtered instruments that regularize persistence while retaining the low-frequency predictive signal that would be lost under first differencing. The construction is conceptually related to the IVX procedures of Kostakis et al. (2015) and Lee (2016), but the expectile setting is not a mechanical extension: the score is smooth but asymmetrically weighted, and IVX instrumentation induces a generally non-symmetric Jacobian, so a separate asymptotic theory is still required.

Our contributions are threefold. First, we develop the first persistence-robust inferential framework for predictive expectile regression. Second, we provide the asymptotic theory for the IVX-expectile estimator, showing that it is consistent and asymptotically normal and that the associated Wald statistic has a standard chi-square limit despite the asymmetric score structure and the generally non-symmetric Jacobian induced by IVX. Third, we provide further empirical evidence to clarify why predictive expectile regression is substantively useful beyond mean and quantile regressions: it captures heterogeneous predictive effects across the conditional distribution and remains sensitive to the magnitude of tail outcomes. In addition, compared with conditional quantile, it admits a direct interpretation in terms of downside tail risk and ES.

Simulation results show that the proposed method controls size well in both univariate and multivariate models across stationary, local-to-unity, and unit root designs, with only a modest local-power cost under very small alternatives and strong power once departures from the null become moderate. In the empirical application to monthly and quarterly U.S. stock return predictability using the updated dataset adopted by Welch and Goyal (2008), we find substantial asymmetry across expectiles, especially in the upper tail, and identify the short rate and default yield spread as the most stable predictors, whereas broader valuation effects are much less stable after 1952.

The remainder of the paper is organized as follows. Section 2 introduces the predictive expectile model and conventional heteroscedasticity and autocorrelation consistent (HAC) inference, together with the developed IVX-expectile procedure. Section 3 presents the asymptotic results. Section 4 reports simulation evidence, and Section 5 studies stock return predictability. We conclude in Section 6. Technical proofs and supplementary simulation results are deferred to Online Appendix.

2 Econometric Model and Methodology

2.1 Mode Setup and Conventional HAC Inference

We begin with a one-step-ahead linear predictive regression for a fixed expectile level $\tau \in (0,1)$:

$$y_{t+1} = \alpha_\tau + x_t'\beta_\tau + u_{t+1,\tau}, \qquad t = 1,\ldots,n-1, \tag{1}$$

where y_{t+1} is the target variable, $x_t \in \mathbb{R}^p$ is a vector of predictors observed at time t, and $u_{t+1,\tau}$ is the expectile error. Let $\mathcal{F}_t := \sigma\{(y_s, x_s) : s \le t\}$ be a sigma-field generated by $\{(y_s, x_s) : s \le t\}$ for each t. Then, the predictive expectile restriction is

$$\mathbb{E}[\psi_\tau(u_{t+1,\tau}) \mid \mathcal{F}_t] = 0, \tag{2}$$

where $\psi_\tau(u) := a_\tau(u)u$ and $a_\tau(u) := \tau\mathbf{1}(u \ge 0) + (1-\tau)\mathbf{1}(u < 0)$, with $\mathbf{1}(\cdot)$ denoting an indicator function. Equation (2) states that the conditional τ-expectile of y_{t+1} given x_t is linear in x_t. Expectiles were introduced by Newey and Powell (1987) as smooth counterparts to regression quantiles (Koenker and Bassett, 1978); see also Efron (1991) for an early interpretation through asymmetric squared error loss. For a nonlinear form of (1), readers are referred to Cai et al. (2018), Cai et al. (2025) and references therein.

The standard estimator follows from asymmetric least squares (ALS). Define

$$\rho_\tau(u) := |\tau - \mathbf{1}(u < 0)|u^2, \tag{3}$$

where $r_t := \begin{pmatrix} 1 \\ x_t \end{pmatrix}$ and $\theta_\tau := \begin{pmatrix} \alpha_\tau \\ \beta_\tau \end{pmatrix}$, and let

$$\hat{\theta}_\tau^{\mathrm{ALS}} = \arg\min_\theta \sum_{t=1}^{n-1} \rho_\tau(y_{t+1} - r_t'\theta)\,, \tag{4}$$

which is the expectile estimate of parameter θ_τ. Because the loss function is differentiable, the estimator can be computed by iteratively reweighted least squares. The first-order condition for (4) is

$$\sum_{t=1}^{n-1} r_t\, \psi_\tau\Big(y_{t+1} - r_t'\hat{\theta}_\tau^{\mathrm{ALS}}\Big) = 0. \tag{5}$$

To describe conventional inference, define the estimating function $g_t(\theta) := r_t\psi_\tau(y_{t+1} - r_t'\theta)$ and the population Jacobian $J_\tau := -\mathbb{E}\Big[\frac{\partial g_t(\theta_\tau)}{\partial \theta'}\Big] = \mathbb{E}[r_t r_t' a_\tau(u_{t+1,\tau})]$. Under standard stationarity, weak-dependence, and moment conditions, $\hat{\theta}_\tau^{\mathrm{ALS}}$ is asymptotically normal with sandwich covariance matrix

$$\sqrt{n}\Big(\hat{\theta}_\tau^{\mathrm{ALS}} - \theta_\tau\Big) \xrightarrow{d} N\Big(0, J_\tau^{-1}\Omega_\tau J_\tau^{-1}\Big)\,, \tag{6}$$

where

$$\Omega_\tau := \sum_{j=-\infty}^{\infty} E[g_t(\theta_\tau) g_{t-j}(\theta_\tau)']\,. \tag{7}$$

Under standard stationary weak-dependence conditions, if the score process $\{g_t(\theta_\tau), \mathcal{F}_t\}$ is covariance stationary, then Ω_τ is the long-run variance of the expectile score process; see, for instance, Cai et al. (2025) for details. If the score sequence is serially correlated, or more generally if one wants an estimator robust to both heteroscedasticity and autocorrelation under stationary weak dependence, inference requires a consistent estimator of Ω_τ. Under the stationary martingale-difference special case, Ω_τ collapses to the contemporaneous covariance and a heteroscedasticity-robust (HC) estimator is asymptotically sufficient. Following Section 2.3 in Cai et al. (2025), a standard kernel type HAC estimator is

$$\hat{\Omega}_\tau^{\mathrm{HAC}} := \hat{\Gamma}_0 + \sum_{j=1}^{L_n} K\Big(\frac{j}{L_n+1}\Big)\Big(\hat{\Gamma}_j + \hat{\Gamma}_j'\Big)\,, \tag{8}$$

where $K(\cdot)$ is a HAC kernel weight function and L_n is the associated bandwidth or truncation lag. In particular, $K(\cdot)$ down-weights higher-order sample auto-covariances as the lag j increases. A standard choice is the Bartlett kernel, $K(u) = (1-|u|)\mathbf{1}(|u| \leq 1)$, although other standard kernels may also be used. Moreover, $\hat{\Gamma}_j := \frac{1}{n}\sum_{t=j+1}^{n-1} \hat{g}_t \hat{g}'_{t-j}$, and $\hat{g}_t := g_t\left(\hat{\theta}_\tau^{\text{ALS}}\right)$. Using the sample Jacobian $\hat{J}_\tau := \frac{1}{n}\sum_{t=1}^{n-1} r_t r'_t a_\tau\left(y_{t+1} - r'_t \hat{\theta}_\tau^{\text{ALS}}\right)$, the conventional HAC covariance estimator becomes

$$\hat{V}_\tau^{\text{HAC}} := \frac{1}{n}\hat{J}_\tau^{-1}\hat{\Omega}_\tau^{\text{HAC}}\hat{J}_\tau^{-1}. \tag{9}$$

For testing linear restrictions $\mathbb{H}_0\colon R\theta_\tau = q$, where R is an $m \times (p+1)$ full-row-rank restriction matrix and $q \in \mathbb{R}^m$ is an unknown constant vector, the corresponding Wald statistic is

$$W_{n,\tau}^{\text{HAC}} := \left(R\hat{\theta}_\tau^{\text{ALS}} - q\right)' \left[R\hat{V}_\tau^{\text{HAC}} R'\right]^{-1} \left(R\hat{\theta}_\tau^{\text{ALS}} - q\right). \tag{10}$$

In stationary applications, one can implement (8)–(10) using standard HAC kernels and bandwidth choices in the spirit of Newey and West (1987) and Andrews (1991); see also Galvao and Yoon (2024) for a recent quantile regression analogue and Cai et al. (2025) for the generalized conditional autoregressive expectile (GARE) setting. This conventional approach underlies much of the applied expectile literature, including dynamic risk-measure models such as Taylor (2008), Kuan et al. (2009), Daouia et al. (2018), Davison et al. (2023), and Cai et al. (2025).

In the literature of financial return predictability, regressors are often highly persistent and may be local to unity, and the innovation in the predictive equation is typically correlated with the innovation driving the predictor. In this case, standard normal approximations and conventional Wald statistics are no longer reliable, even in mean regression, as emphasized by Campbell and Yogo (2006), Mikusheva (2007), Kostakis et al. (2015), and Phillips and Lee (2016). A similar theoretical challenge arises in predictive quantile regressions with persistent covariates, where valid inference requires methods adapted to persistence, endogeneity, and heteroscedasticity; see, among others, Lee (2016), Fan and Lee (2019), and Cai et al. (2023). By contrast, the theory for predictive expectile regression under persistent predictors remains largely undeveloped. This gap motivates the IVX-based expectile procedure developed in the next section.

2.2 IVX-Expectile

The idea of IVX, pioneered by Magdalinos and Phillips (2009) and extended by Kostakis et al. (2015) and Lee (2016), among others, is to replace the original persistent regressor with an instrument that is less persistent than x_t, but still retains substantially more low-frequency information than first differencing. This construction delivers a feasible instrument that is robust to local-to-unity persistence while preserving nontrivial power in predictive regressions. We adapt this idea to predictive expectile regression by combining the IVX filter with the expectile score in Section 2.1.

For a p-dimensional predictor, let $\Delta x_t := x_t - x_{t-1}$, where $x_t = R_{nx} x_{t-1} + v_t$ with $R_{nx} := I_p + C/n$, and construct the IVX-filtered regressor recursively as

$$\tilde{z}_t = R_{nz} \tilde{z}_{t-1} + \Delta x_t, \quad \text{where} \quad R_{nz} := I_p + C_z/n^{\delta} \tag{11}$$

with $\tilde{z}_0 = 0$ and $\delta \in (0,1)$, I_p is a $p \times p$ identity matrix, and $C_z = c_z I_p$ with $c_z < 0$, so that the scalar persistence parameter c_z controls how quickly the filter damps the low-frequency variation in x_t. We then define the IVX instrument vector

$$z_t = (1,\ \tilde{z}_t')'. \tag{12}$$

Compared with the original regressor $r_t = (1, x_t')'$, the IVX instrument z_t is asymptotically less persistent and hence better behaved for inference, yet it avoids the severe loss of signal that would result from using only Δx_t. Throughout this study, the coordinates of x_t are treated within the common local-to-unity framework in (11). This allows differential persistence across predictors, including exact unit-root and highly persistent stationary cases (i.e., c_z is large but finite).

Given z_t, define the IVX-expectile moment function

$$g_n^{\mathrm{IVX}}(\theta) := \frac{1}{n-1}\sum_{t=1}^{n-1} z_t \psi_\tau(y_{t+1} - r_t'\theta)\,. \tag{13}$$

Our baseline estimator is the just-identified IVX-expectile estimator

$$\hat{\theta}_\tau^{\mathrm{IVX}} \in \left\{\theta : g_n^{\mathrm{IVX}}(\theta) = 0\right\}. \tag{14}$$

The root in (14) is computed numerically using the iteration described below. In over-identified specifications, one may instead minimize the standard generalized method of moment (GMM) criterion

$$Q_n(\theta) := g_n^{\mathrm{IVX}}(\theta)' W_n g_n^{\mathrm{IVX}}(\theta),$$

for a positive definite weighting matrix W_n. In either case, the central modification relative to ALS is that the regression score is instrumented by z_t rather than by the original regressor r_t, which changes both the effective persistence of the estimating equations and the resulting covariance structure.

For applied researchers, the IVX-expectile procedure can be summarized as follows.

Step 1. Fix τ and compute the ALS estimator $\hat{\theta}_\tau^{\mathrm{ALS}}$ from (4).

Step 2. Select the pair of (δ, C_z) in (11). Following the IVX literature, δ should be close to, but strictly smaller than, one, while C_z should be negative definite. In empirical work, it is common to use a scalar specification $C_z = c_z I_p$ with $c_z < 0$. Throughout our simulations and empirical application we use the default choice $\delta = 0.95$ and $c_z = -5$, and then check robustness over nearby values of (δ, c_z) when needed.

Step 3. Compute $\tilde{z}_t$ recursively from (11) and form z_t as in (12).

Step 4. Starting from $\theta^{(0)} = \hat{\theta}_\tau^{\text{ALS}}$, iterate on

$$\theta^{(m+1)} = \theta^{(m)} + \left[\hat{J}_n^{\text{IVX}}\left(\theta^{(m)}\right)\right]^{-1} g_n^{\text{IVX}}\left(\theta^{(m)}\right),$$

where

$$\hat{J}_n^{\text{IVX}}(\theta) := \frac{1}{n-1}\sum_{t=1}^{n-1} z_t r_t' a_\tau(y_{t+1} - r_t'\theta)$$

is the sample Jacobian in the sense that $\partial g_n^{\text{IVX}}(\theta)/\partial\theta' = -\hat{J}_n^{\text{IVX}}(\theta)$. Under the local non-singularity conditions stated in Section 3, this Newton step is locally well behaved when initialized at the consistent ALS estimator. The iteration stops once successive updates are sufficiently small, yielding $\hat{\theta}_\tau^{\text{IVX}}$.

Remark 1. *We do not separately develop an optimal, data-driven selector for* (δ, C_z) *in this study. Our goal is instead to establish validity over the IVX framework and to use a conventional benchmark implementation in finite samples. In the related IVX-QR framework, Lee (2016) provides a formal discussion of the choice of* (δ, C_z) *and suggests a practical selection rule based on the size-power tradeoff. We do not adapt that quantile-specific procedure to the present expectile setting. Accordingly, we adopt the common scalar specification* $C_z = c_z I_p$ *with* $c_z < 0$ *and use* $(\delta, c_z) = (0.95, -5)$ *as the benchmark choice throughout, and nearby values can be adopted as robustness checks. Developing a formal tuning-parameter selection rule for IVX-expectile regression is an interesting extension that merits a future study.*

Inference proceeds exactly as in Section 2.1, except that the score is instrumented by z_t instead of r_t. Let $g_t^{\text{IVX}}(\theta) := z_t\psi_\tau(y_{t+1} - r_t'\theta)$ and $\hat{g}_t^{\text{IVX}} := g_t^{\text{IVX}}\left(\hat{\theta}_\tau^{\text{IVX}}\right)$. Under Assumption 2 in Section 3, we estimate the covariance matrix of the IVX score process using the same HAC form as in Section 2.1. Under the maintained martingale-difference benchmark, it nests the HC special case asymptotically because the off-diagonal autocovariances vanish.

The long-run variance of the IVX score process is then estimated by

$$\hat{\Omega}_\tau^{\mathrm{IVX}} := \hat{\Gamma}_0^{\mathrm{IVX}} + \sum_{j=1}^{L_n} K\left(\frac{j}{L_n+1}\right)\left(\hat{\Gamma}_j^{\mathrm{IVX}} + \hat{\Gamma}_j^{\mathrm{IVX}\prime}\right), \tag{15}$$

where $\hat{\Gamma}_j^{\mathrm{IVX}} := \frac{1}{n}\sum_{t=j+1}^{n-1} \hat{g}_t^{\mathrm{IVX}} \hat{g}_{t-j}^{\mathrm{IVX}\prime}$, with the kernel $K(\cdot)$ and bandwidth L_n defined as in Section 2.1. Using the Jacobian estimate $\hat{J}_\tau^{\mathrm{IVX}} := \hat{J}_n^{\mathrm{IVX}}\left(\hat{\theta}_\tau^{\mathrm{IVX}}\right)$, the IVX sandwich covariance matrix is

$$\hat{V}_\tau^{\mathrm{IVX}} := \frac{1}{n}\left(\hat{J}_\tau^{\mathrm{IVX}}\right)^{-1} \hat{\Omega}_\tau^{\mathrm{IVX}} \left(\hat{J}_\tau^{\mathrm{IVX}}\right)^{-1\prime}. \tag{16}$$

Unlike the ALS covariance matrix in Section 2.1, the rightmost factor is the transpose inverse because $J_\tau^{\mathrm{IVX}} = \mathbb{E}[z_t r_t' a_\tau(u_{t+1,\tau})]$ is generally not symmetric once the score is instrumented. For testing linear restrictions $\mathbb{H}_0\colon\ R\theta_\tau = q$, where R is an $m\times(p+1)$ full-row-rank restriction matrix and $q \in \mathbb{R}^m$, we use the IVX-Wald statistic

$$W_{n,\tau}^{\mathrm{IVX}} := \left(R\hat{\theta}_\tau^{\mathrm{IVX}} - q\right)' \left[R\hat{V}_\tau^{\mathrm{IVX}} R'\right]^{-1} \left(R\hat{\theta}_\tau^{\mathrm{IVX}} - q\right) \tag{17}$$

to report results of the IVX-Wald tests and confidence intervals.

Although the construction in this section is deliberately parallel to the ALS setup in Section 2.1, the IVX-expectile problem is not a purely mechanical substitution of z_t for r_t. The expectile score remains smooth but τ-asymmetric, the Jacobian is weighted by the sign-dependent factor $a_\tau(\cdot)$, and once the score is instrumented by z_t the resulting Jacobian is generally not symmetric. These features matter for both computation and inference, and the next section shows that they can nonetheless be handled within a persistence-robust IVX asymptotic framework.

3 Asymptotic Theory

This section is devoted to establishing the asymptotic results for the IVX-expectile procedure. Throughout, we fix $\tau \in (0,1)$ and consider the following setup:

$$y_{t+1} = \alpha_\tau + x_t'\beta_\tau + u_{t+1,\tau} \quad \text{with} \quad x_t = R_{nx}x_{t-1} + v_t \quad \text{and} \quad R_{nx} := I_p + C/n, \tag{18}$$

where C is a fixed $p \times p$ matrix. This setup nests highly persistent stationary predictors and the local-to-unity case emphasized in predictive regressions (Campbell and Yogo, 2006; Mikusheva, 2007; Phillips and Magdalinos, 2007). Let $\tilde{z}_t$ be the IVX filter in (11), $z_t = (1, \tilde{z}_t')'$, and define the normalization matrices

$$H_n := \operatorname{diag}\Big(1, n^{-\delta/2} I_p\Big) , \quad S_n := \sqrt{n}\, H_n = \operatorname{diag}\Big(n^{1/2}, n^{(1-\delta)/2} I_p\Big) , \quad D_n := \operatorname{diag}\Big(n^{1/2}, n I_p\Big) . \tag{19}$$

The normalized IVX score is

$$\bar{g}_{t,n}^{\mathrm{IVX}}(\theta) := H_n z_t \psi_\tau(y_{t+1} - r_t'\theta) , \tag{20}$$

so that

$$S_n g_n^{\mathrm{IVX}}(\theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n-1} \bar{g}_{t,n}^{\mathrm{IVX}}(\theta).$$

Although the proof strategy for these asymptotic results follows the established IVX literature for predictive mean and quantile regression, the predictive expectile setting differs in two respects that require separate arguments. First, the score $\psi_\tau(\cdot)$ is continuous and Lipschitz but its derivative changes discontinuously at zero through the weight $a_\tau(\cdot)$, so local linearization requires a no-mass-at-zero and conditional-density regularity condition. Second, once the score is instrumented by z_t, the Jacobian involves $z_t r_t' a_\tau(u_{t+1,\tau})$ and is generally asymmetric. The target of this section is to show that these expectile-specific features are still compatible with the same persistence-robust IVX normalization that yields normal and Wald limits.

Assumption 1. *The predictor follows (18) with $x_0 = O_p(1)$, and the IVX filter is constructed as in (11) with $\delta \in (0,1)$ and a fixed matrix C_z whose eigenvalues have strictly negative real parts.*

Assumption 2. *For each t, the regressor vectors r_t and z_t are $\mathcal{F}_t$-measurable, while $(u_{t+1,\tau}, v_{t+1}')'$ is $\mathcal{F}_{t+1}$-measurable. Moreover:*

(i) $\mathbb{E}[\psi_\tau(u_{t+1,\tau}) \mid \mathcal{F}_t] = 0$ *almost surely;*

(ii) $\{\bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau), \mathcal{F}_t\}$ *is a martingale-difference sequence;*

(iii) for some $\eta > 0$, $\sup_{n,t} \mathbb{E}[\|\bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau)\|^{2+\eta}] < \infty$*;*

(iv) $u_{t+1,\tau}$ *has conditional density continuous in a neighborhood of zero and* $\Pr(u_{t+1,\tau} = 0 \mid \mathcal{F}_t) = 0$ *almost surely;*

(v) there exists a positive definite matrix $\Omega_\tau^{\mathrm{IVX}}$ *such that*

$$\frac{1}{n}\sum_{t=1}^{n-1} \mathbb{E}\Big[\bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau)\bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau)' \mid \mathcal{F}_t\Big] \ \xrightarrow{p}\ \Omega_\tau^{\mathrm{IVX}}. \tag{21}$$

Assumption 3. *There exists a compact neighborhood* $\mathcal{N}_\tau$ *of* θ_τ *such that* θ_τ *is the unique solution of the population IVX moment restriction and* $\sup_{\theta \in \mathcal{N}_\tau} \left\| S_n \hat{J}_n^{\mathrm{IVX}}(\theta) D_n^{-1} - J_\tau^{\mathrm{IVX}} \right\| \xrightarrow{p} 0$ *for some nonsingular matrix* J_τ^{IVX}.

Assumption 4. *The kernel* $K(\cdot)$ *is even, bounded, and continuous at the origin, with* $K(0) = 1$ *and compact support. The bandwidth satisfies* $L_n \to \infty$ *and* $L_n/\sqrt{n} \to 0$.

These assumptions are standard in spirit for IVX-based predictive inference, but a few deserve further discussion. Assumption 1 imposes the usual local-to-unity persistence together with the mildly integrated IVX filter. In Assumption 2, (i) is the population expectile restriction, and because H_n is deterministic and z_t is $\mathcal{F}_t$-measurable, it implies the martingale-difference property in (ii). We state (ii) separately to keep the theory close to the existing IVX arguments but narrows the dependence class relative to more general mixing setups. Accordingly, the key dependence condition is a martingale-difference structure for the normalized IVX score, and we retain the HAC covariance form throughout, both to provide a single implementation formula and as a connection to extensions that accommodate serial dependence in the score. (iv) is the key expectile-specific regularity condition. Since $\psi_\tau(\cdot)$ is differentiable except for a kink in its derivative at zero, continuity of the conditional density near zero and the absence of a point mass there ensure that local mean-value expansions remain valid around θ_τ. (v) is a condition that drives the martingale central limit theorem in Lemma 1. As in the analogous IVX-quantile setting (Cai et al., 2023), its finite-sample performance depends on how well the conditional variance average approximates its limit. Our simulation results in Section 4 suggest that this concern is less pronounced in the present expectile framework, where size control remains accurate even at moderate sample size. Assumption 3 is a local identification condition on the normalized IVX Jacobian. As in the IVX literature, one can derive it from more primitive restrictions on the joint law of $(x_t, u_{t+1,\tau})$, but we state it directly to keep the focus on the persistence-robust normalization rather than on a separate Jacobian derivation. Primitive sufficient conditions can be verified by combining the near-unit-root growth bounds for (x_t, z_t) with standard uniform laws of large numbers for the weighted Jacobian array, but spelling them out would add notation without changing the main IVX argument. Finally, Assumption 4 keeps the feasible covariance estimator in HAC form. Under the maintained martingale-difference benchmark, the higher-order autocovariance terms are asymptotically negligible, so an HC estimator would suffice asymptotically. We retain the HAC form because it matches the feasible estimator used in practice and remains valid under mild deviations from exact martingale-difference behavior. Extending the theory to broader weak dependence is left for future work.

Lemma 1. *Suppose Assumptions 1 and 2 hold. Then we have* $\max_{1\le t\le n}\|x_t\| = O_p(n^{1/2})$, $\max_{1\le t\le n}\|\tilde{z}_t\| = O_p(n^{\delta/2})$, *and* $\frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}\bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau) \xrightarrow{d} N\left(0, \Omega_\tau^{\mathrm{IVX}}\right)$.

Lemma 1 formalizes the key IVX mechanism. The persistent predictor grows at the local-to-unity rate $n^{1/2}$, whereas the IVX instrument grows only at the milder rate $n^{\delta/2}$. Hence the normalized score in (20) is asymptotically tight, which is precisely what restores standard normal inference. The growth orders follow the same logic as in Magdalinos and Phillips

(2009), Kostakis et al. (2015), and Lee (2016), while the normal limit follows from martingale arguments as in Hall and Heyde (2014) and Hansen (1992).

Lemma 2. *Suppose Assumptions 1–4 hold. Then* $S_n \hat{J}_\tau^{\mathrm{IVX}} D_n^{-1} \xrightarrow{p} J_\tau^{\mathrm{IVX}}$*, and the HAC estimator consistently estimates the long-run variance of the normalized IVX score* $H_n \hat{\Omega}_\tau^{\mathrm{IVX}} H_n' \xrightarrow{p} \Omega_\tau^{\mathrm{IVX}}$.

Lemma 2 shows that the IVX Jacobian and the IVX-HAC covariance estimator remain well behaved after the natural normalizations in (19). This is the direct expectile analogue of the Jacobian and long-run variance regularity conditions used in IVX mean and quantile predictive regressions.

Theorem 1. *Suppose Assumptions 1–4 hold. Then there exists a sequence of solutions* $\hat{\theta}_\tau^{\mathrm{IVX}}$ *to (14) such that* $\hat{\theta}_\tau^{\mathrm{IVX}} \xrightarrow{p} \theta_\tau$*, and* $D_n\left(\hat{\theta}_\tau^{\mathrm{IVX}} - \theta_\tau\right) = \left(J_\tau^{\mathrm{IVX}}\right)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n-1} \bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau) + o_p(1)$*. Consequently, we have* $\hat{\alpha}_\tau^{\mathrm{IVX}} - \alpha_\tau = O_p(n^{-1/2})$ *and* $\hat{\beta}_\tau^{\mathrm{IVX}} - \beta_\tau = O_p(n^{-1})$.

Remark 2. *Theorem 1 implies that the intercept retains the usual root-n rate, while the slope vector is estimated at rate n. The faster slope rate reflects the near-integrated nature of the predictor, whereas IVX removes the endogeneity-driven nonstandardity that would otherwise contaminate the limiting distribution.*

Theorem 2. *Suppose Assumptions 1–4 hold. Then we have*

$$D_n\left(\hat{\theta}_\tau^{\mathrm{IVX}} - \theta_\tau\right) \xrightarrow{d} N\left(0, V_\tau^{\mathrm{IVX}}\right), \quad \text{where} \quad V_\tau^{\mathrm{IVX}} := \left(J_\tau^{\mathrm{IVX}}\right)^{-1} \Omega_\tau^{\mathrm{IVX}} \left(J_\tau^{\mathrm{IVX}}\right)^{-1\prime}.$$

Moreover, we have $D_n \hat{V}_\tau^{\mathrm{IVX}} D_n \xrightarrow{p} V_\tau^{\mathrm{IVX}}$.

Theorem 2 is the central asymptotic justification for the proposed estimator. After applying the IVX normalization, the estimator is asymptotically normal with covariance consistently estimated by the IVX-HAC sandwich matrix. Thus, the nuisance effects of persistence and endogeneity are absorbed by the instrument transformation and the long-run variance estimator rather than by the limiting law itself. The normalization matrix D_n is only a theoretical device: it is not computed in practice, and it cancels from the feasible Wald statistic in Theorem 3.

Theorem 3. *Suppose Assumptions 1–4 hold and let R be a fixed $m \times (p+1)$ full-row-rank matrix and $q \in \mathbb{R}^m$. Under the null hypothesis $\mathbb{H}_0\colon R\theta_\tau = q$, then,*

$$W_{n,\tau}^{\mathrm{IVX}} \xrightarrow{d} \chi_m^2.$$

Hence, the IVX-Wald test that rejects for large values of $W_{n,\tau}^{\mathrm{IVX}}$ is asymptotically of correct size, and the corresponding confidence region has asymptotic coverage $1-\alpha$.

Remark 3. *Theorem 3 explains why IVX is attractive in predictive expectile regressions. Conventional ALS-HAC inference is valid under standard stationary weak dependence, but it generally fails to deliver a pivotal limit when predictors are nearly integrated and endogenous. By contrast, the IVX transformation restores a standard χ^2 limit without requiring the practitioner to estimate nuisance persistence parameters directly, just as in the mean and quantile predictive-regression settings studied by Kostakis et al. (2015) and Lee (2016).*

Remark 4. *For notational convenience, the preceding results are stated for the just-identified IVX-expectile estimator in Section 2.2. The same proof strategy extends to overidentified IVX-expectile GMM estimators, provided the additional instruments satisfy analogues of Assumptions 2–4 and the corresponding GMM weighting matrix is consistently estimated.*

The formal proofs of Lemmas 1–2 and Theorems 1–3 are deferred to Online Appendix A.

4 Numerical Studies

This section evaluates the finite-sample performance of the proposed IVX-expectile estimator through a comprehensive Monte Carlo study. We start with a univariate predictive expectile regression, and then consider a multivariate model with three predictors.

4.1 Univariate Model

The baseline univariate data generating process (DGP) is formulated as $y_{t+1} = \alpha_\tau + \beta_\tau x_t + u_{t+1,\tau}$, where the predictor follows an AR(1) process $x_t = \rho_n\, x_{t-1} + v_t$. Persistence is governed by the local-to-unity parameterization $\rho_n = 1 - c/n$, and we vary $c \in \{0, 2, 5, 10, 25, 50\}$ so that smaller c implies stronger persistence. The embedded endogeneity is introduced by correlating $u_{t+1,\tau}$ and v_{t+1} with a fixed correlation of -0.95. For each expectile level τ, the innovation $u_{t+1,\tau}$ is generated so that its conditional τ-expectile is zero, preserving a linear conditional expectile structure. Regarding the regression disturbance, we consider the following scenarios:

(i) Joint Normal. The innovations $(u_{t+1,\tau}, v_{t+1})$ are jointly normal with mean $(0,0)'$ and correlation matrix $(1, -0.95; -0.95, 1)$.

(ii) Joint Student's t. The normal expectile disturbance in (i) is replaced by a standardized $t(5)$ distribution, recentered so that its marginal τ-expectile is zero before forming $u_{t+1,\tau}$.

(iii) GARCH Volatility. The innovation takes the form $u_{t+1,\tau} = \sigma_{t+1} e_{t+1,\tau}$ with $\sigma_{t+1}^2 = 0.05 + 0.05\, u_{t,\tau}^2 + 0.9\, \sigma_t^2$, where $e_t \sim$ i.i.d.$N(0,1)$.

Throughout this section, we let sample size $n \in \{250, 500, 1000\}$ and the expectile level $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$. Choice of the pair of (δ, C_z) in (11) is given at STEP 2 in Section 2.2. Each simulation is repeated 2000 times. We compare IVX-expectile with the naive expectile regression based on conventional Wald inference, which ignores predictor persistence and endogeneity, as well as the HAC-based Wald inference defined in (10).

Size performance. Table 1 reports rejection frequencies under the null hypothesis $\mathbb{H}_0\colon \beta_\tau = 0$ at the nominal 5% level under the normal disturbance scenario. IVX-expectile exhibits excellent size control over the full grid of (n, c, τ) combinations, including tail expectiles and the unit root case $c = 0$. By contrast, the naive and HAC procedures substantially over-reject when the predictor is highly persistent, especially for $c \in \{0, 2, 5\}$. Their size distortion becomes less severe as persistence declines, but it remains uniformly less reliable than the IVX-based inference. Similar patterns can be observed from the other two disturbance scenarios, as shown by Appendix Table A1.

Table 1: Empirical rejection frequencies under H_0: Univariate model with normal disturbance

		$\tau = 0.1$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.9$		
T	c	naive	HAC	IVX	naive	HAC	IVX	naive	HAC	IVX	naive	HAC	IVX	naive	HAC	IVX
250	0	0.2140	0.2295	0.0635	0.2475	0.2690	0.0580	0.2840	0.3040	0.0505	0.2595	0.2770	0.0635	0.2260	0.2400	0.0505
	2	0.1575	0.1670	0.0615	0.1755	0.1930	0.0595	0.1800	0.2045	0.0560	0.1575	0.1735	0.0550	0.1525	0.1625	0.0555
	5	0.1135	0.1220	0.0565	0.1075	0.1210	0.0535	0.1085	0.1230	0.0515	0.1170	0.1315	0.0520	0.1075	0.1270	0.0660
	10	0.0850	0.0940	0.0635	0.1040	0.1145	0.0600	0.1035	0.1145	0.0595	0.0935	0.0995	0.0490	0.0900	0.0970	0.0615
	25	0.0815	0.0940	0.0700	0.0800	0.0885	0.0645	0.0750	0.0855	0.0480	0.0565	0.0680	0.0510	0.0785	0.0835	0.0640
	50	0.0660	0.0745	0.0605	0.0535	0.0610	0.0515	0.0610	0.0700	0.0470	0.0695	0.0780	0.0600	0.0695	0.0795	0.0680
500	0	0.2190	0.2320	0.0415	0.2625	0.2755	0.0585	0.2885	0.2960	0.0605	0.2340	0.2500	0.0610	0.2080	0.2205	0.0450
	2	0.1375	0.1460	0.0515	0.1495	0.1620	0.0540	0.1730	0.1850	0.0510	0.1570	0.1725	0.0535	0.1335	0.1395	0.0505
	5	0.1020	0.1130	0.0595	0.1180	0.1260	0.0630	0.1100	0.1240	0.0560	0.1185	0.1255	0.0455	0.1125	0.1145	0.0630
	10	0.0805	0.0870	0.0560	0.0980	0.1070	0.0490	0.1000	0.1025	0.0510	0.0970	0.1065	0.0495	0.0900	0.0930	0.0510
	25	0.0645	0.0695	0.0575	0.0665	0.0790	0.0560	0.0675	0.0770	0.0495	0.0690	0.0740	0.0560	0.0695	0.0775	0.0575
	50	0.0560	0.0625	0.0480	0.0665	0.0700	0.0515	0.0670	0.0745	0.0580	0.0630	0.0700	0.0565	0.0615	0.0630	0.0505
1000	0	0.2180	0.2280	0.0560	0.2545	0.2670	0.0490	0.2765	0.2855	0.0570	0.2715	0.2780	0.0525	0.2010	0.2140	0.0555
	2	0.1380	0.1425	0.0545	0.1585	0.1615	0.0630	0.1680	0.1755	0.0470	0.1590	0.1630	0.0415	0.1220	0.1270	0.0385
	5	0.0955	0.0975	0.0535	0.1020	0.1100	0.0540	0.1075	0.1135	0.0565	0.1095	0.1170	0.0445	0.1055	0.1095	0.0560
	10	0.0785	0.0800	0.0460	0.0725	0.0790	0.0430	0.0760	0.0805	0.0525	0.0800	0.0865	0.0515	0.0885	0.0945	0.0590
	25	0.0655	0.0670	0.0545	0.0600	0.0695	0.0440	0.0635	0.0605	0.0450	0.0645	0.0680	0.0470	0.0620	0.0645	0.0510
	50	0.0535	0.0590	0.0470	0.0595	0.0670	0.0510	0.0570	0.0635	0.0460	0.0500	0.0595	0.0455	0.0525	0.0535	0.0485

Size-adjusted power. Next, we compare the local power performance of the three methods in each disturbance scenario. To trace out power curves, we fix $n = 1000$ and vary the slope coefficient over $\beta \in \{0.000, 0.005, 0.010, \ldots, 0.095, 1.000\}$, so that $\beta = 0$ provides the null benchmark. The expectile grid, persistence grid, and endogeneity correlation are the same as in the size study. To make power comparisons comparable across methods, we report size-adjusted power: for each combination of (c, τ) and each method, we first estimate the finite-sample critical value from the simulated null distribution at $\beta = 0$, and then apply that critical value to the alternative values of β. The resulting power curves therefore compare different settings on an equal size-adjusted basis.

Figure 1: Comparison of size-adjusted power: Univariate model with normal disturbance

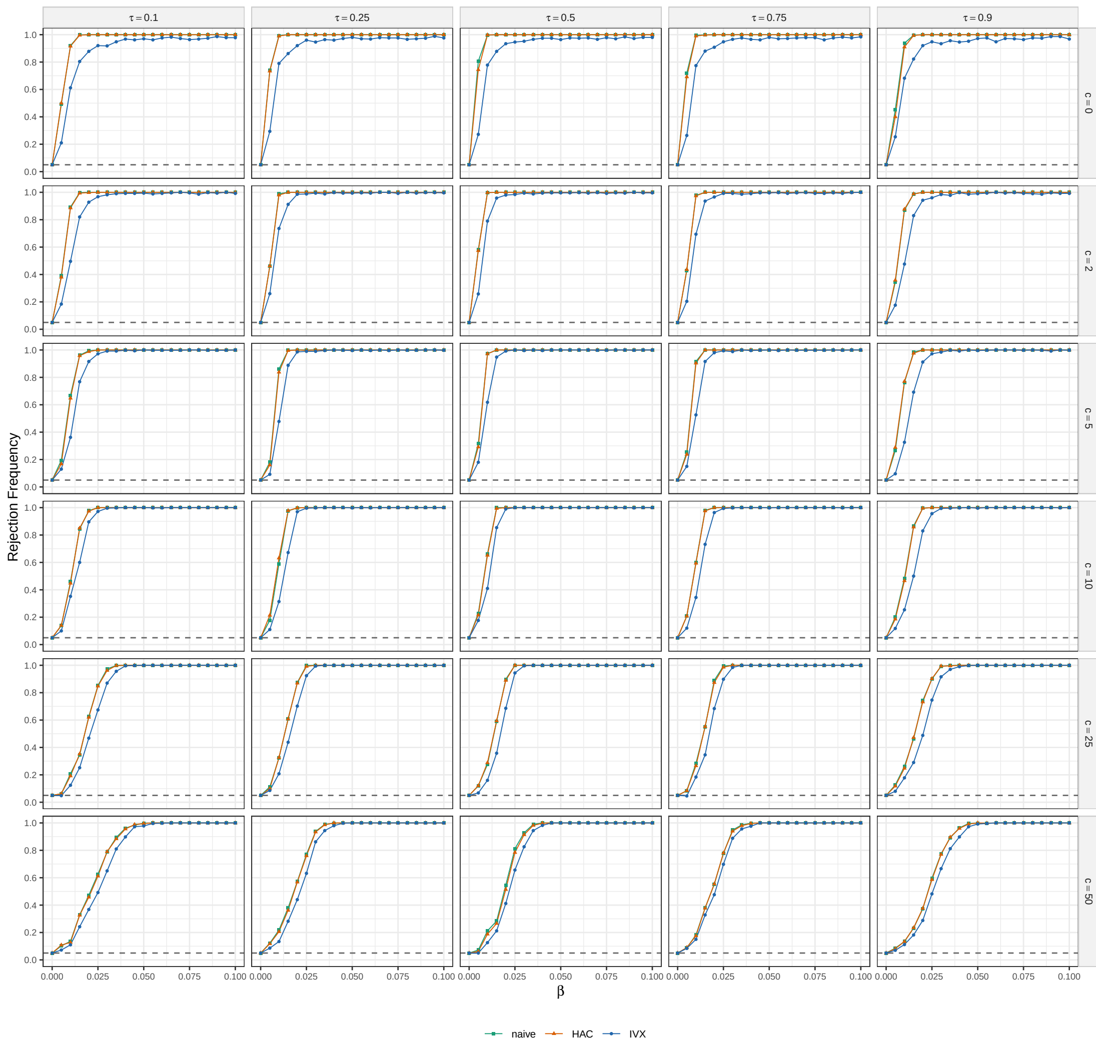


Figure 1 presents the size-adjusted power results under the normal disturbance scenario. It reveals three main observations. First, IVX-expectile does pay a local-power cost relative to the naive and HAC procedures, but this cost is concentrated at very small alternatives. For example, when $\beta_\tau = 0.005$, the IVX power is about 0.15, compared with approximately 0.28 for the size-adjusted naive and HAC methods, whose power curves trace each other closely and are broadly indistinguishable. However, the gap between the three power curves shrinks quickly as the slope coefficient grows: by $\beta_\tau = 0.05$, the average IVX power broadly approaches 1 across the pair of (c, τ). Second, all three methods' power is strongest around the middle expectiles and weakest at the tails. Third, for a fixed alternative, power is highest when the predictor is most persistent and declines as persistence weakens. For example, the average IVX power at $\beta_\tau = 0.01$ falls from about 0.73 when $c = 0$ to 0.13 when $c = 50$. Similar patterns can be observed for the other two disturbance scenarios, as displayed by Appendix Figures A1 and A2.

4.2 Multivariate Model

We next consider a three-predictor expectile regression formulated as $y_{t+1} = \alpha_\tau + x_t'\beta_\tau + u_{t+1,\tau}$ with $x_t = (x_{1t}, x_{2t}, x_{3t})'$, and evaluate size under the joint null $\beta_\tau = (0,0,0)'$. Each predictor follows $x_{j,t} = \rho_{j,n} x_{j,t-1} + v_{j,t}$ with $\rho_{j,n} = 1 - c_j/n$ and $(c_1, c_2, c_3) = (0, 5, 50)$, so the design combines a unit-root regressor, a local-to-unity regressor, and a weakly persistent regressor. To introduce endogeneity while keeping the predictors only mildly correlated, we draw $(u_{t+1,\tau}, v_{1,t+1}, v_{2,t+1}, v_{3,t+1})'$ from a centered multivariate normal distribution with mean zero, variances one, $\mathrm{Corr}(u_{t+1,\tau}, v_{j,t+1}) = -0.6$ for each j, and $\mathrm{Corr}(v_{i,t+1}, v_{j,t+1}) = 0.2$ for $i \neq j$. The sample size $n \in \{250, 500, 1000\}$ and the expectile level $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ are the same as in the univariate model, and each design point is replicated 2000 times.

Panels A to C in Table 2 report the empirical rejection frequencies of the individual tests $\mathbb{H}_0\colon \beta_{j,\tau} = 0$ for $j = 1, 2, 3$, and panel D reports the results of the joint test $\mathbb{H}_0\colon \beta_{1,\tau} = \beta_{2,\tau} = \beta_{3,\tau} = 0$. The same qualitative pattern as in the univariate model continues to hold, but the multivariate setting makes the size distortion even more visible in the joint test. The naive procedure still substantially over-rejects for all three coefficients, especially for the unit-root regressor, and the conventional HAC correction does not meaningfully alleviate this distortion even as n increases. By contrast, the IVX procedure remains much closer to nominal size. Although it is still slightly oversized in the tails for the joint test, size control improves markedly once the sample size reaches 1000.

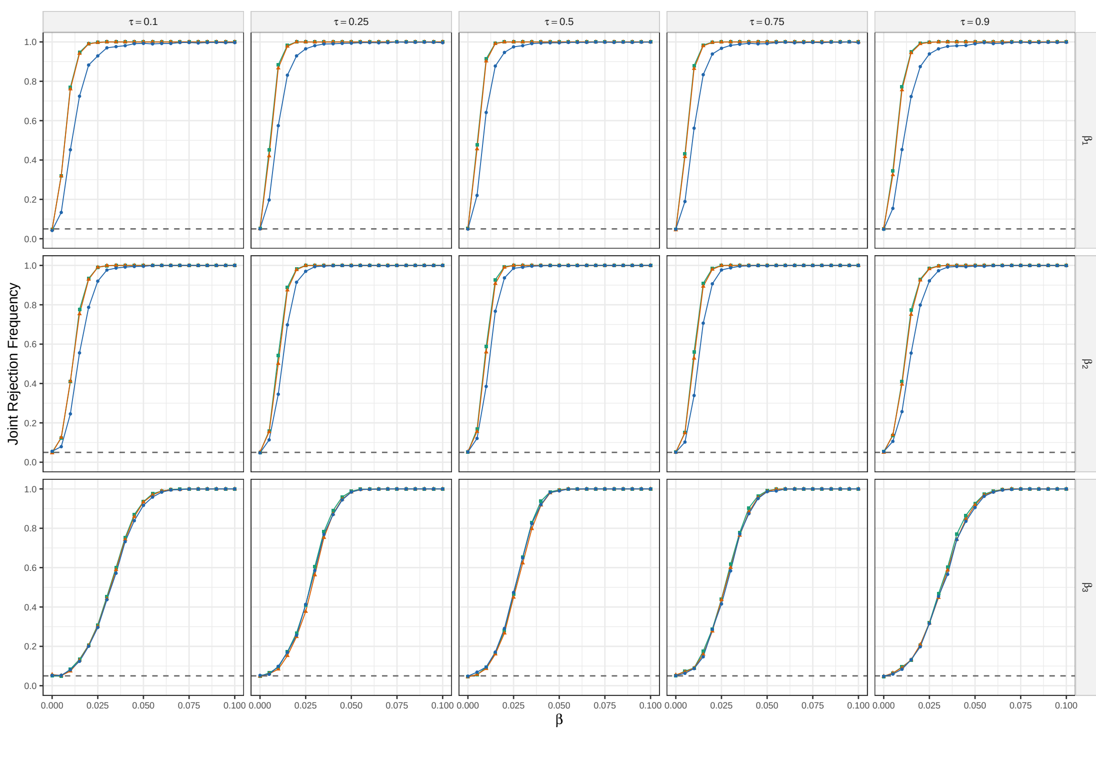
Table 2: Empirical rejection frequencies under the null: Multivariate model

	$\tau = 0.1$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.9$		
T	naive	HAC	IVX	naive	HAC	IVX	naive	HAC	IVX	naive	HAC	IVX	naive	HAC	IVX
Panel A. Unit-root regressor ($H_0\colon \beta_{1,\tau} = 0$)															
250	0.1575	0.1645	0.0580	0.1825	0.1995	0.0625	0.2015	0.2070	0.0675	0.1840	0.1965	0.0725	0.1560	0.1680	0.0615
500	0.1425	0.1510	0.0550	0.1775	0.1905	0.0605	0.1865	0.2005	0.0505	0.1705	0.1760	0.0545	0.1625	0.1815	0.0670
1000	0.1485	0.1550	0.0515	0.1850	0.1920	0.0560	0.1775	0.1850	0.0550	0.1815	0.1805	0.0595	0.1400	0.1480	0.0525
Panel B. Local-to-unity regressor ($H_0\colon \beta_{2,\tau} = 0$)															
250	0.1195	0.1300	0.0835	0.1140	0.1220	0.0595	0.1280	0.1445	0.0820	0.1275	0.1430	0.0695	0.1180	0.1220	0.0790
500	0.1040	0.1155	0.0565	0.1240	0.1345	0.0640	0.1310	0.1460	0.0685	0.1205	0.1230	0.0735	0.1135	0.1210	0.0645
1000	0.1035	0.1095	0.0595	0.1255	0.1355	0.0675	0.1215	0.1320	0.0670	0.1205	0.1215	0.0675	0.0940	0.0965	0.0575
Panel C. Weakly persistent regressor ($H_0\colon \beta_{3,\tau} = 0$)															
250	0.0795	0.0785	0.0630	0.0660	0.0840	0.0580	0.0725	0.0810	0.0625	0.0755	0.0830	0.0630	0.0780	0.0845	0.0725
500	0.0725	0.0795	0.0675	0.0705	0.0740	0.0590	0.0650	0.0700	0.0505	0.0765	0.0845	0.0645	0.0780	0.0880	0.0685
1000	0.0665	0.0730	0.0610	0.0665	0.0715	0.0555	0.0695	0.0730	0.0525	0.0675	0.0690	0.0520	0.0655	0.0675	0.0525
Panel D. Joint test ($H_0\colon \beta_{1,\tau} = \beta_{2,\tau} = \beta_{3,\tau} = 0$)															
250	0.1610	0.1895	0.0850	0.1665	0.2000	0.0625	0.1735	0.2185	0.0750	0.1715	0.2050	0.0745	0.1635	0.1845	0.0910
500	0.1390	0.1580	0.0645	0.1505	0.1680	0.0630	0.1525	0.1720	0.0575	0.1475	0.1680	0.0635	0.1560	0.1775	0.0795
1000	0.1250	0.1430	0.0555	0.1445	0.1595	0.0590	0.1410	0.1555	0.0610	0.1415	0.1540	0.0605	0.1215	0.1395	0.0550

We next examine the joint-test power in the multivariate design by fixing $n = 1000$ and increasing one slope coefficient at a time while holding the other two at zero. Figure 2 reports the resulting size-adjusted rejection frequencies of the joint test. For example, the first row of Figure 2 denotes the rejection frequencies for the null hypothesis $\mathbb{H}_0\colon \beta_{1,\tau} = \beta_{2,\tau} = \beta_{3,\tau} = 0$ as we increase $\beta_{1,\tau}$ from 0 to 0.1 while keeping $\beta_{2,\tau}$ and $\beta_{3,\tau}$ at zero.

Results in Figure 2 reveal patterns similar to those in the univariate model. First, IVX again pays a local-power cost relative to the naive and HAC procedures, but the gap is concentrated at very small alternatives. However, IVX power rises quickly as the corresponding slope coefficient grows. Second, the power ordering mirrors the persistence ordering of the regressors. The joint test is most sensitive when the coefficient on the unit-root regressor moves away from zero ($\beta_{1,\tau}$), next most sensitive for the local-to-unity regressor ($\beta_{2,\tau}$), and weakest when only the weakly persistent regressor is active ($\beta_{3,\tau}$). For example, at $\beta_{j,\tau} = 0.01$, the average IVX rejection rate declines from around 0.5 along the β_1 path to around 0.3 along the β_2 path, and to only around 0.1 along the β_3 path. This ranking is expected because the more persistent regressors generate stronger low-frequency signal, which the IVX filter is designed to preserve. Third, as in the univariate power study, power is strongest around the middle expectiles and weaker in the tails. In sum, Figure 2 shows that IVX remains slightly conservative locally, but still delivers strong power once the departure from the null is moderate, especially for the more persistent predictors.

Figure 2: Comparison of size-adjusted power: Multivariate model



5 Predictability of Stock Returns

We apply the proposed IVX-expectile procedure to assess the predictive content of several widely used financial and macroeconomic variables for excess stock market returns. The sample runs from 1927 to 2024 at both monthly and quarterly frequencies, and the dataset is an updated version of that used by Welch and Goyal (2008).[1] Specifically, we use S&P 500 value-weighted log excess returns as our measure of excess market returns and consider the following predictors: dividend payout ratio (`d/e`), long-term yield (`lty`), dividend yield (`d/y`), dividend-price ratio (`d/p`), T-bill rate (`tbl`), earnings-price ratio (`e/p`), book-to-market value ratio (`b/m`), default yield spread (`dfy`), net equity expansion (`ntis`), term spread (`tms`), inflation rate (`inf`), and consumption-wealth ratio (`cay`), which is available only at quarterly frequency from 1952 onward.

Appendix Table A2 reports the AR(1) coefficients together with augmented Dickey-Fuller (Said and Dickey, 1984) and Phillips-Perron (Phillips and Perron, 1988) test statistics for both the full sample (1927–2024) and the post-1952 subsample, separately for monthly (panel A) and quarterly (panel B) data. This subsample is of particular interest in the literature because term-structure variables are often viewed as more informative after the Federal Reserve policy shift emphasized by Kostakis et al. (2015), while predictability in the more recent sample is often thought to have weakened with improved market efficiency (Campbell and Yogo, 2006). We draw two main conclusions. First, persistence is substantial at both frequencies. For monthly data, most predictors have AR(1) coefficients above 0.96, with the long-term yield, dividend yield, dividend-price ratio, and Treasury bill rate especially close to unity in both sample windows. Quarterly persistence is slightly lower on average, but these same predictors remain highly persistent, with autoregressive coefficients still between about 0.94 and 0.98. Second, the yield and valuation predictors provide the

[1] The dataset is publicly available at https://sites.google.com/view/agoyal145.

clearest evidence of near-unit-root behavior. In both monthly samples, the ADF and PP tests fail to reject a unit root for `lty`, `d/y`, and `d/p`, while `tbl` is only marginally rejected by the ADF test but not by the PP test. The same result largely carries over to the quarterly data, where `lty`, `d/y`, `d/p`, and `tbl` all fail to reject in both sample windows. By contrast, spreads and financing variables such as `dfy`, `ntis`, and `tms`, together with inflation and the earnings-price ratio, more clearly reject a unit root despite remaining persistent, and the quarterly consumption-wealth ratio also rejects once it becomes available after 1952. Although the unit root tests cannot reach agreement on every predictor regarding the presence of unit root, several of the empirically important predictors are highly persistent and often statistically indistinguishable from unit-root processes. This provides direct empirical motivation for the IVX-based predictive expectile regressions that follow.

We apply the IVX-expectile procedure to each predictor in both univariate and multivariate models, using the IVX filter parameters $\delta = 0.95$ and $c_z = -5$ and the expectile grid $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ throughout. Panel A in Table 3 reports the monthly IVX-expectile estimates, with the full-sample results for 1927–2024 shown in the left block of columns and the post-1952 results shown in the right block. The full-sample monthly estimates reveal pronounced asymmetry. One clear example is `dfy`: its slope coefficient is strongly negative at $\tau = 0.10$, close to zero at $\tau = 0.50$, and strongly positive at $\tau = 0.75$ and $\tau = 0.90$. This pattern can be interpreted as a widening credit spread worsening downside outcomes while simultaneously increasing upper-tail return potential. The standard valuation ratios load mainly on the upper tail: both `d/y` and `d/p` are significant only at $\tau = 0.75$ and $\tau = 0.90$, while `b/m` is strongest at the upper expectiles but also shows weak lower-tail significance at $\tau = 0.10$. The dividend-earnings ratio `d/e` is similarly asymmetric, with negative coefficients in the lower tail and positive coefficients in the upper tail. Among the interest-rate predictors, `tbl` is significantly negative from the median upward, `lty` is weakly significant around the middle and upper-middle expectiles, and `tms` changes sign across the distribution, being marginally negative at $\tau = 0.10$ but positive in the upper tail. By contrast, `e/p` matters only in the lower tail, `ntis` enters only through weak negative upper-tail effects, and `inf` is significant only around $\tau = 0.75$.

Table 3: Predictability tests by IVX-expectile: Univariate model

Panel A. Monthly

	Full sample: 1927–2024					Subsample: 1952–2024				
Predictor	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
`d/e`	-0.0458***	-0.0217*	-0.0007	0.0184*	0.0471**	-0.0082	-0.0003	0.0047	0.0077	0.0100*
	(0.0168)	(0.0125)	(0.0092)	(0.0104)	(0.0229)	(0.0130)	(0.0101)	(0.0073)	(0.0057)	(0.0054)
`lty`	-0.0159	-0.0901	-0.1241*	-0.1211*	-0.1262	-0.1032	-0.1357	-0.1386*	-0.1036	-0.0668
	(0.1071)	(0.0773)	(0.0664)	(0.0709)	(0.0996)	(0.1185)	(0.0857)	(0.0744)	(0.0795)	(0.1078)
`d/y`	-0.0118	0.0008	0.0120	0.0251**	0.0423***	0.0026	0.0044	0.0058	0.0080	0.0113**
	(0.0114)	(0.0076)	(0.0073)	(0.0105)	(0.0161)	(0.0089)	(0.0063)	(0.0051)	(0.0050)	(0.0057)
`d/p`	-0.0177	-0.0030	0.0099	0.0242***	0.0421***	-0.0010	0.0027	0.0057	0.0095*	0.0144**
	(0.0111)	(0.0076)	(0.0065)	(0.0090)	(0.0141)	(0.0088)	(0.0063)	(0.0051)	(0.0052)	(0.0062)
`tbl`	0.1062	-0.0228	-0.1073*	-0.1770**	-0.2868**	-0.1168	-0.1397*	-0.1352**	-0.1110*	-0.0960
	(0.1049)	(0.0766)	(0.0634)	(0.0751)	(0.1141)	(0.1061)	(0.0773)	(0.0630)	(0.0613)	(0.0785)
`e/p`	0.0168*	0.0111*	0.0079	0.0063	0.0064	0.0043	0.0022	0.0013	0.0014	0.0027
	(0.0087)	(0.0065)	(0.0053)	(0.0056)	(0.0075)	(0.0094)	(0.0063)	(0.0054)	(0.0057)	(0.0072)
`b/m`	-0.0372*	-0.0080	0.0199	0.0531**	0.0936***	-0.0046	-0.0015	0.0035	0.0116	0.0207*
	(0.0192)	(0.0141)	(0.0146)	(0.0212)	(0.0286)	(0.0166)	(0.0109)	(0.0088)	(0.0092)	(0.0116)
`dfy`	-2.4138***	-1.1245*	0.2804	1.8063**	3.4485***	-0.9407	-0.4197	0.1802	0.8283*	1.4561***
	(0.5640)	(0.6146)	(0.6919)	(0.8844)	(1.0373)	(0.7998)	(0.6763)	(0.5690)	(0.4959)	(0.5118)
`ntis`	0.1351	-0.0093	-0.1484	-0.3143*	-0.5413*	0.1108	-0.0046	-0.0491	-0.0945	-0.1565
	(0.2279)	(0.1505)	(0.1267)	(0.1848)	(0.2941)	(0.2521)	(0.1745)	(0.1257)	(0.1162)	(0.1374)
`tms`	-0.4617*	-0.1435	0.0697	0.3154*	0.7029**	0.1958	0.1587	0.1160	0.1148	0.1519
	(0.2666)	(0.1786)	(0.1385)	(0.1862)	(0.3265)	(0.2298)	(0.1419)	(0.1115)	(0.1072)	(0.1311)
`inf`	0.2765	-0.2062	-0.5221	-0.6434*	-0.7878	-1.1973	-1.0931*	-1.0174**	-0.8275*	-0.6713
	(0.5514)	(0.4245)	(0.3367)	(0.3482)	(0.5095)	(0.8993)	(0.6380)	(0.5062)	(0.4606)	(0.5295)

Panel B. Quarterly

	Full sample: 1927–2024					Subsample: 1952–2024				
Predictor	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
`d/e`	-0.0742*	-0.0343	0.0105	0.0537	0.1458	-0.0017	0.0110	0.0209	0.0257**	0.0255***
	(0.0436)	(0.0327)	(0.0265)	(0.0480)	(0.0929)	(0.0256)	(0.0219)	(0.0159)	(0.0117)	(0.0079)
`lty`	-0.1404	-0.2961	-0.3333*	-0.3283*	-0.4196*	-0.2826	-0.3850	-0.3603	-0.2840	-0.3049
	(0.3334)	(0.2397)	(0.1961)	(0.1960)	(0.2220)	(0.3629)	(0.2703)	(0.2209)	(0.2099)	(0.2157)
`d/y`	-0.0033	0.0142	0.0369**	0.0653***	0.1060**	0.0238	0.0238	0.0256	0.0281	0.0273
	(0.0266)	(0.0200)	(0.0170)	(0.0250)	(0.0417)	(0.0248)	(0.0191)	(0.0173)	(0.0176)	(0.0230)
`d/p`	-0.0122	0.0125	0.0448*	0.0838**	0.1362***	0.0137	0.0160	0.0217	0.0280*	0.0320
	(0.0292)	(0.0222)	(0.0248)	(0.0382)	(0.0524)	(0.0246)	(0.0198)	(0.0170)	(0.0168)	(0.0203)
`tbl`	0.1454	-0.1146	-0.3027	-0.5068**	-0.8769**	-0.2013	-0.3390	-0.3551*	-0.3453**	-0.3914**
	(0.3062)	(0.2306)	(0.1924)	(0.2372)	(0.4087)	(0.3302)	(0.2383)	(0.1914)	(0.1724)	(0.1717)
`e/p`	0.0389	0.0328	0.0273	0.0274	0.0328	0.0116	0.0058	0.0027	0.0003	-0.0055
	(0.0249)	(0.0200)	(0.0181)	(0.0203)	(0.0259)	(0.0231)	(0.0181)	(0.0156)	(0.0154)	(0.0166)
`b/m`	-0.0201	0.0239	0.0963*	0.1847**	0.2845***	0.0013	0.0036	0.0171	0.0321	0.0410
	(0.0415)	(0.0382)	(0.0536)	(0.0792)	(0.0859)	(0.0477)	(0.0326)	(0.0265)	(0.0261)	(0.0378)
`dfy`	-3.5549**	-1.4233	1.8500	5.9608**	10.4438***	-1.4838	-0.8680	0.6784	2.2703*	3.1535***
	(1.8021)	(1.5793)	(2.2217)	(3.0275)	(3.2160)	(1.2802)	(1.4909)	(1.5147)	(1.2412)	(1.2053)
`ntis`	-0.1708	-0.3639	-0.7120*	-1.2577	-2.0390	0.2103	0.0087	-0.1681	-0.2691	-0.3511
	(0.4099)	(0.3539)	(0.4192)	(0.8051)	(1.3869)	(0.5807)	(0.4825)	(0.3979)	(0.3481)	(0.3749)
`tms`	-0.8178	-0.3023	0.2300	0.9165	1.9521	0.0268	0.2376	0.3022	0.3926	0.4721
	(0.6295)	(0.4956)	(0.4356)	(0.6656)	(1.2392)	(0.5816)	(0.4243)	(0.3507)	(0.3418)	(0.3535)
`inf`	0.9580	0.1711	-0.6812	-1.6106	-2.7851*	-1.1595	-1.2029	-1.0548	-0.7946	-0.5906
	(1.0178)	(0.8783)	(0.8215)	(1.1486)	(1.5311)	(1.0277)	(0.8129)	(0.6578)	(0.5886)	(0.5756)
`cay`	–	–	–	–	–	0.1906	0.1055	0.0252	-0.0259	-0.0151
	–	–	–	–	–	(0.3499)	(0.2793)	(0.2097)	(0.1397)	(0.0820)

The post-1952 monthly estimates are much more selective. Lower-tail significance largely disappears, but the upper-tail effects of `dfy` remain clearly visible, especially at $\tau = 0.90$, and the upper-tail coefficients on `d/p`, `d/y`, `d/e`, and `b/m` remain significant at conventional levels. Meanwhile, short-rate variables become relatively more important than valuation ratios in the middle portion of the distribution: `tbl` is significant at $\tau = 0.25$, $\tau = 0.50$, and $\tau = 0.75$, and `inf` is significantly negative over the same range, with `lty` remaining marginally significant at $\tau = 0.50$. In contrast, `tms`, `e/p`, and `ntis` lose significance completely. The monthly evidence therefore suggests that part of the broader full-sample predictability is tied to the pre-1952 period, a finding consistent with Campbell and Yogo (2006) and Kostakis et al. (2015), while the post-war sample preserves a narrower combination of upper-tail valuation effects and middle-to-upper-tail short-rate effects.

Panel B presents the same comparison for quarterly data. The full-sample quarterly results again show stronger and broader evidence than the post-1952 subsample. In the full quarterly sample, aggregation sharpens the medium- and upper-expectile effects of the valuation variables: `d/y`, `d/p`, and `b/m` are all significant from $\tau = 0.50$ upward, with the largest coefficients at $\tau = 0.90$. The default yield spread continues to display the same tail asymmetry seen in the monthly data, remaining negative at $\tau = 0.10$, statistically insignificant near the middle, and strongly positive at $\tau = 0.75$ and $\tau = 0.90$. The short rate `tbl` is significant only in the upper tail, while `lty` is weakly negative from the median upward. By contrast, `tms` is insignificant throughout, `ntis` is only marginally significant at the median, and `e/p` remains insignificant throughout. The post-1952 quarterly results are more selective. The most robust effects are the positive upper-tail coefficients on `dfy` and the negative middle-to-upper-tail coefficients on `tbl`. In addition, `d/e` becomes significantly positive only at $\tau = 0.75$ and $\tau = 0.90$, while `d/p` is only marginally significant at $\tau = 0.75$. The broader valuation pattern present in the full quarterly sample largely disappears in the post-1952 sample: `d/y` and `b/m` are no longer significant, and neither `lty`, `inf`, nor `ntis` exhibits reliable predictive ability. The `cay`, available only from 1952 onward, is insignificant at all studied expectile levels.

Table 4: Predictability tests by IVX-expectile: Multivariate model with monthly data

	Full sample: 1927–2024								Subsample: 1952–2024							
Expectile	`d/p`	`tbl`	`dfy`	`tms`	`b/m`	`d/e`	`e/p`	Joint Wald	`d/p`	`tbl`	`dfy`	`tms`	`b/m`	`d/e`	`e/p`	Joint Wald
Panel A. Ang and Bekaert (2007)																
$\tau = 0.10$	-0.0187	0.1309	–	–	–	–	–	3.054	0.0010	-0.1192	–	–	–	–	–	1.211
$\tau = 0.25$	-0.0028	-0.0173	–	–	–	–	–	0.290	0.0057	-0.1563*	–	–	–	–	–	3.489
$\tau = 0.50$	0.0115*	-0.1301*	–	–	–	–	–	4.307	0.0087	-0.1591**	–	–	–	–	–	6.535**
$\tau = 0.75$	0.0267***	-0.2281***	–	–	–	–	–	9.944***	0.0121**	-0.1448**	–	–	–	–	–	9.163**
$\tau = 0.90$	0.0453***	-0.3687***	–	–	–	–	–	12.136***	0.0170***	-0.1484**	–	–	–	–	–	11.513***
Panel B. Ferson and Schadt (1996)																
$\tau = 0.10$	0.0031	0.0718	-2.5539***	0.1463	–	–	–	19.219***	0.0036	-0.0352	-1.1101	0.2063	–	–	–	5.453
$\tau = 0.25$	0.0080	-0.0516	-1.2876*	-0.0074	–	–	–	6.272	0.0069	-0.1351	-0.3853	0.0297	–	–	–	8.123*
$\tau = 0.50$	0.0116**	-0.1790**	0.1535	-0.1701	–	–	–	9.078*	0.0082	-0.2084*	0.3878	-0.1199	–	–	–	7.266
$\tau = 0.75$	0.0156***	-0.2887***	1.6855*	-0.3226**	–	–	–	18.466***	0.0096*	-0.2431***	1.0806*	-0.2077	–	–	–	9.729**
$\tau = 0.90$	0.0238***	-0.4371***	3.2753***	-0.5510***	–	–	–	32.142***	0.0127**	-0.2699***	1.5845**	-0.2624	–	–	–	15.618***
Panel C. Kothari and Shanken (1997)																
$\tau = 0.10$	0.0026	–	–	–	-0.0402	–	–	3.692	0.0032	–	–	–	-0.0093	–	–	0.100
$\tau = 0.25$	0.0028	–	–	–	-0.0111	–	–	0.344	0.0102	–	–	–	-0.0158	–	–	0.669
$\tau = 0.50$	-0.0015	–	–	–	0.0216	–	–	2.383	0.0123	–	–	–	-0.0137	–	–	1.750
$\tau = 0.75$	-0.0094	–	–	–	0.0635*	–	–	10.316***	0.0124	–	–	–	-0.0059	–	–	3.610
$\tau = 0.90$	-0.0205	–	–	–	0.1161***	–	–	18.867***	0.0144	–	–	–	0.0001	–	–	5.787*
Panel D. Lamont (1998)																
$\tau = 0.10$	-0.0042	–	–	–	–	-0.0441***	–	7.623**	0.0013	–	–	–	–	-0.0087	–	0.400
$\tau = 0.25$	0.0028	–	–	–	–	-0.0226*	–	3.247	0.0030	–	–	–	–	-0.0012	–	0.216
$\tau = 0.50$	0.0109*	–	–	–	–	-0.0040	–	3.625	0.0050	–	–	–	–	0.0033	–	1.333
$\tau = 0.75$	0.0211***	–	–	–	–	0.0114	–	7.505**	0.0081	–	–	–	–	0.0055	–	4.103
$\tau = 0.90$	0.0328***	–	–	–	–	0.0328*	–	9.700***	0.0126*	–	–	–	–	0.0062	–	6.960**
Panel E. Campbell and Vuolteenaho (2004)																
$\tau = 0.10$	–	–	–	-0.0362	-0.0607***	–	0.0330***	11.251**	–	–	–	0.2519	-0.0265	–	0.0157	3.279
$\tau = 0.25$	–	–	–	0.0284	-0.0241	–	0.0188*	3.576	–	–	–	0.1876	-0.0129	–	0.0081	2.700
$\tau = 0.50$	–	–	–	0.0364	0.0171	–	0.0028	3.827	–	–	–	0.1239	0.0014	–	0.0016	1.483
$\tau = 0.75$	–	–	–	0.0506	0.0629**	–	-0.0123	8.765**	–	–	–	0.0999	0.0155	–	-0.0029	2.635
$\tau = 0.90$	–	–	–	0.0560	0.1135***	–	-0.0275**	16.647***	–	–	–	0.1170	0.0268*	–	-0.0048	4.644

Following Kostakis et al. (2015), we additionally extend the univariate analysis to multivariate specifications by considering five predictor groups studied in the earlier literature: (1) `d/p` and `tbl` (Ang and Bekaert, 2007), (2) `d/p`, `tbl`, `dfy`, and `tms` (Ferson and Schadt, 1996), (3) `d/p` and `b/m` (Kothari and Shanken, 1997), (4) `d/p` and `d/e` (Lamont, 1998), and (5) `e/p`, `tms`, and `b/m` (Campbell and Vuolteenaho, 2004). Results in Table 4 reveal two features. First, the groups that include the short rate remain the most robust. In both the full sample and the post-1952 subsample, the pair (`d/p`, `tbl`) is jointly significant in the upper half of the expectile range, with a positive loading on `d/p` and a negative loading on `tbl`. The broader group (`d/p`, `tbl`, `dfy`, `tms`) is even stronger: in the full sample it is jointly significant at $\tau = 0.10$, $\tau = 0.50$, $\tau = 0.75$, and $\tau = 0.90$, and in the post-1952 sample at $\tau = 0.25$, $\tau = 0.75$, and $\tau = 0.90$. Within that specification, `dfy` preserves the same lower- and upper-tail sign reversal seen in the univariate results, while `tbl` remains sharply negative from the middle to the upper tail. Second, the valuation-only combinations are much less stable. The groups (`d/p`, `b/m`) and (`d/p`, `d/e`) are jointly significant mainly in the full-sample upper tail, but most of that significance disappears after 1952. The group (`e/p`, `b/m`, `tms`) is jointly significant in the full sample, with `e/p` loading positively in the lower tail and negatively in the upper tail while `b/m` shifts from negative to positive, but these effects largely disappear in the post-war subsample. The monthly multivariate evidence therefore supports the univariate finding that post-war predictability is concentrated in short-rate and credit-spread variables rather than in broad valuation bundles.

Table 5 shows a similar pattern at quarterly frequency. The pair (`d/p`, `tbl`) is jointly significant from $\tau = 0.50$ upward in both samples, and the larger group (`d/p`, `tbl`, `dfy`, `tms`) is again the strongest specification, with joint significance in the full sample at $\tau = 0.50$, $\tau = 0.75$, and $\tau = 0.90$, and in the post-1952 sample at $\tau = 0.75$ and $\tau = 0.90$. The short-rate coefficient remains strongly negative in the upper tail, while `dfy` becomes strongly positive there and sharply negative at $\tau = 0.10$ in the full sample. By contrast, the specifications centered on `b/m` and `d/e` are more sample-dependent. The group (`d/p`, `b/m`) is jointly significant only in the full sample upper tail, whereas (`d/p`, `d/e`) remains jointly significant in the upper tail even after 1952, although more weakly. The group (`e/p`, `b/m`, `tms`) also becomes jointly significant in the upper tail of the full quarterly sample but only marginally so after 1952, driven mainly by a positive loading on `b/m` and a negative loading on `e/p`.

The overall empirical evidence from the univariate and multivariate analyses on both monthly and quarterly frequencies indicates three primary conclusions. First, stock return predictability is remarkably asymmetric: many predictors are largely insignificant at the conditional middle but begin to exhibit meaningful predictive power in the tails, especially the upper tail. Second, the most stable predictive ability comes from the short rate and the default yield spread: `tbl` is repeatedly negative from the middle to the upper tail, while `dfy` continues to display a recurring sign reversal across tails in both univariate and multivariate specifications. Third, subsample stability is limited for broad valuation ratios and valuation-only combinations, whereas short-rate and credit-spread variables retain their predictive ability in the post-war period. These features would be difficult to detect in a standard mean-regression analysis, and therefore illustrate the value of combining expectile methods with persistence-robust IVX inference.

Table 5: Predictability tests by IVX-expectile: Multivariate model with quarterly data

	Full sample: 1927–2024								Subsample: 1952–2024							
Expectile	`d/p`	`tbl`	`dfy`	`tms`	`b/m`	`d/e`	`e/p`	Joint Wald	`d/p`	`tbl`	`dfy`	`tms`	`b/m`	`d/e`	`e/p`	Joint Wald
Panel A. Ang and Bekaert (2007)																
$\tau = 0.10$	-0.0133	0.1620	–	–	–	–	–	0.352	0.0176	-0.2439	–	–	–	–	–	0.862
$\tau = 0.25$	0.0138	-0.1349	–	–	–	–	–	0.483	0.0224	-0.3914	–	–	–	–	–	3.041
$\tau = 0.50$	0.0497*	-0.4032*	–	–	–	–	–	4.851*	0.0297	-0.4382**	–	–	–	–	–	6.105**
$\tau = 0.75$	0.0901**	-0.6901***	–	–	–	–	–	7.563**	0.0363**	-0.4530***	–	–	–	–	–	9.479***
$\tau = 0.90$	0.1441***	-1.1492***	–	–	–	–	–	9.684***	0.0421**	-0.5192***	–	–	–	–	–	14.909***
Panel B. Ferson and Schadt (1996)																
$\tau = 0.10$	0.0252	-0.0649	-4.0507**	-0.0643	–	–	–	4.675	0.0262	-0.2864	-1.6558	-0.2332	–	–	–	3.392
$\tau = 0.25$	0.0349*	-0.3346	-1.8740	-0.4499	–	–	–	5.888	0.0271	-0.4400	-0.7029	-0.2391	–	–	–	5.025
$\tau = 0.50$	0.0433**	-0.5802**	1.4063	-0.6941	–	–	–	9.437*	0.0281	-0.5813*	1.1728	-0.3501	–	–	–	6.175
$\tau = 0.75$	0.0521***	-0.8739***	5.4436*	-1.0251*	–	–	–	16.002***	0.0282	-0.6633**	2.8757	-0.4644	–	–	–	9.391*
$\tau = 0.90$	0.0664***	-1.3510***	9.4951***	-1.5332***	–	–	–	33.523***	0.0279	-0.8313***	4.0243**	-0.7005	–	–	–	11.907**
Panel C. Kothari and Shanken (1997)																
$\tau = 0.10$	-0.0036	–	–	–	-0.0165	–	–	0.234	0.0373	–	–	–	-0.0540	–	–	0.685
$\tau = 0.25$	0.0001	–	–	–	0.0238	–	–	0.396	0.0420	–	–	–	-0.0568	–	–	1.338
$\tau = 0.50$	-0.0148	–	–	–	0.1125	–	–	4.086	0.0406	–	–	–	-0.0398	–	–	2.089
$\tau = 0.75$	-0.0460	–	–	–	0.2357*	–	–	9.560***	0.0379	–	–	–	-0.0201	–	–	2.913
$\tau = 0.90$	-0.0789	–	–	–	0.3710***	–	–	21.653***	0.0368	–	–	–	-0.0101	–	–	2.628
Panel D. Lamont (1998)																
$\tau = 0.10$	0.0116	–	–	–	–	-0.0789*	–	2.985	0.0154	–	–	–	–	-0.0065	–	0.337
$\tau = 0.25$	0.0248	–	–	–	–	-0.0430	–	2.584	0.0144	–	–	–	–	0.0066	–	0.729
$\tau = 0.50$	0.0457**	–	–	–	–	-0.0035	–	3.973	0.0179	–	–	–	–	0.0159	–	2.486
$\tau = 0.75$	0.0760**	–	–	–	–	0.0276	–	5.525*	0.0227	–	–	–	–	0.0199	–	5.715*
$\tau = 0.90$	0.1099***	–	–	–	–	0.0868	–	8.756**	0.0251	–	–	–	–	0.0201**	–	10.663***
Panel E. Campbell and Vuolteenaho (2004)																
$\tau = 0.10$	–	–	–	-0.3034	-0.0447	–	0.0490	3.169	–	–	–	0.1318	-0.0412	–	0.0270	0.698
$\tau = 0.25$	–	–	–	-0.1312	0.0020	–	0.0312	2.716	–	–	–	0.2831	-0.0141	–	0.0126	0.680
$\tau = 0.50$	–	–	–	-0.0658	0.1007	–	-0.0044	4.229	–	–	–	0.2958	0.0217	–	-0.0030	1.192
$\tau = 0.75$	–	–	–	-0.0128	0.2182**	–	-0.0396	7.981**	–	–	–	0.3066	0.0470	–	-0.0126	3.248
$\tau = 0.90$	–	–	–	0.0302	0.3446***	–	-0.0825*	17.641***	–	–	–	0.3285	0.0597	–	-0.0175	7.204*

Finally, the lower-tail expectile results also admit a direct interpretation in terms of ES. Newey and Powell (1987) establish a one-to-one mapping between expectile levels and quantile levels, and Bellini et al. (2014) show that each expectile coincides with the ES at the corresponding quantile level, making expectiles coherent risk measures that are also elicitable (Ziegel, 2016). This elicitability property means that the conditional ES can be estimated directly via the asymmetric least squares loss in (3), circumventing the well-known difficulty that ES is not elicitable on its own (Gneiting, 2011). In our application, the negative `dfy` coefficient at $\tau = 0.10$ in the full monthly and quarterly samples implies that a widening credit spread shifts the lower conditional expectile downward, corresponding to worse downside tail outcomes for investors. Conversely, lower-tail evidence for the valuation variables is weak and unstable: `d/y` and `d/p` are never significant at $\tau = 0.10$, `b/m` is only marginally significant in the lower tail of the full monthly sample, and `d/e` shows lower-tail significance in the full monthly sample and only weakly in the full quarterly sample, but not in the post-1952 results. Therefore, the IVX-expectile framework does more than identify asymmetric predictability: it also helps isolate which predictors contain robust predictive information about downside tail risk once persistence is taken into account.

6 Conclusion

This paper develops an IVX-based framework for predictive expectile regressions with highly persistent predictors. By combining expectile score equations with IVX instruments, the proposed procedure preserves the distributional perspective of expectile regression while delivering persistence-robust inference in settings where conventional HAC-based methods can be unreliable because of near-unit-root behavior, endogeneity, and heteroscedasticity. We establish consistency and asymptotic normality for the IVX-expectile estimator and show that the associated Wald statistic has a standard chi-square limit. Our simulation results indicate that the method controls size well across stationary, local-to-unity, and unit-root designs, with only a modest local-power cost relative to naive procedures under very small alternatives. In the empirical analysis of stock return predictability, the IVX-expectile approach detects substantial asymmetry across expectiles, highlights the relatively stable predictive roles of the short rate and default yield spread, and shows that broader valuation-based predictability is much less stable across sample periods. These findings show that IVX-expectile regression provides a useful and conservative tool for studying heterogeneous predictive effects and downside tail behavior when covariates are persistent.

Disclosure statement

The authors report there are no competing interests to declare.

References

Andrews, D. W. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* pp. 817–858.

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

Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E. (2014), 'Generalized quantiles as risk measures', *Insurance: Mathematics and Economics* **54**, 41–48.

Cai, Z., Chen, H. and Liao, X. (2023), 'A new robust inference for predictive quantile regression', *Journal of Econometrics* **234**(1), 227–250.

Cai, Z., Fang, Y. and Tian, D. (2018), 'Assessing tail risk using expectile regressions with partially varying coefficients', *Journal of Management Science and Engineering* **3**(4), 183–213.

Cai, Z., Fang, Y. and Tian, D. (2025), 'Assessing tail risk via a generalized conditional autoregressive expectile model', *Journal of Financial Econometrics* **23**(2), nbaf010.

Campbell, J. Y. and Shiller, R. J. (1988), 'The dividend-price ratio and expectations of future dividends and discount factors', *Review of Financial Studies* **1**(3), 195–228.

Campbell, J. Y. and Vuolteenaho, T. (2004), 'Bad beta, good beta', *American Economic Review* **94**(5), 1249–1275.

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

Daouia, A., Girard, S. and Stupfler, G. (2018), 'Estimation of tail risk based on extreme expectiles', *Journal of the Royal Statistical Society Series B* **80**(2), 263–292.

Davison, A. C., Padoan, S. A. and Stupfler, G. (2023), 'Tail risk inference via expectiles in heavy-tailed time series', *Journal of Business & Economic Statistics* **41**(3), 876–889.

Efron, B. (1991), 'Regression percentiles using asymmetric squared error loss', *Statistica Sinica* **1**(1), 93–125.

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

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

Galvao, A. F. and Yoon, J. (2024), 'Hac covariance matrix estimation in quantile regression', *Journal of the American Statistical Association* **119**(547), 2305–2316.

Gneiting, T. (2011), 'Making and evaluating point forecasts', *Journal of the American Statistical Association* **106**(494), 746–762.

Hall, P. and Heyde, C. C. (2014), *Martingale limit theory and its application*, Academic press.

Hansen, B. E. (1992), 'Convergence to stochastic integrals for dependent heterogeneous processes', *Econometric Theory* **8**(4), 489–500.

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

Koenker, R. and Bassett, G. (1978), 'Regression quantiles', *Econometrica* pp. 33–50.

Koenker, R. and Xiao, Z. (2004), 'Unit root quantile autoregression inference', *Journal of the American Statistical Association* **99**(467), 775–787.

Koenker, R. and Xiao, Z. (2006), 'Quantile autoregression', *Journal of the American Statistical Association* **101**(475), 980–990.

Kostakis, A., Magdalinos, T. and Stamatogiannis, M. P. (2015), 'Robust econometric inference for stock return predictability', *Review of Financial Studies* **28**(5), 1506–1553.

Kothari, S. P. and Shanken, J. (1997), 'Book-to-market, dividend yield, and expected market returns: A time-series analysis', *Journal of Financial Economics* **44**(2), 169–203.

Kuan, C.-M., Yeh, J.-H. and Hsu, Y.-C. (2009), 'Assessing value at risk with CARE, the conditional autoregressive expectile models', *Journal of Econometrics* **150**(2), 261–270.

Lamont, O. (1998), 'Earnings and expected returns', *Journal of Finance* **53**(5), 1563–1587.

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

Magdalinos, T. and Phillips, P. C. (2009), 'Econometric inference in the vicinity of unity', *Unpublished working paper* .

Maynard, A., Shimotsu, K. and Kuriyama, N. (2024), 'Inference in predictive quantile regressions', *Journal of Econometrics* **245**(1-2), 105875.

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

Newey, W. K. and Powell, J. L. (1987), 'Asymmetric least squares estimation and testing', *Econometrica* **55**(4), 819–847.

Newey, W. and West, K. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

Pesaran, M. H. and Timmermann, A. (1995), 'Predictability of stock returns: Robustness and economic significance', *Journal of Finance* **50**(4), 1201–1228.

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

Phillips, P. C. and Magdalinos, T. (2007), 'Limit theory for moderate deviations from a unit root', *Journal of Econometrics* **136**(1), 115–130.

Phillips, P. C. and Perron, P. (1988), 'Testing for a unit root in time series regression', *Biometrika* **75**(2), 335–346.

Said, S. E. and Dickey, D. A. (1984), 'Testing for unit roots in autoregressive-moving average models of unknown order', *Biometrika* **71**(3), 599–607.

Stambaugh, R. F. (1999), 'Predictive regressions', *Journal of Financial Economics* **54**(3), 375–421.

Taylor, J. W. (2008), 'Estimating value at risk and expected shortfall using expectiles', *Journal of Financial Econometrics* **6**(2), 231–252.

Tian, D., Cai, Z. and Fang, Y. (2019), 'Econometric modeling of risk measures: A selective review of the recent literature', *Applied Mathematics-A Journal of Chinese Universities* **34**(2), 205–228.

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

Ziegel, J. F. (2016), 'Coherence and elicitability', *Mathematical Finance* **26**(4), 901–918.

SUPPLEMENTARY MATERIAL

A Technical Proofs

Proof of Lemma 1. Define the predictor recursion in stacked form as

$$x_t = R_{nx}^t x_0 + \sum_{j=1}^{t} R_{nx}^{t-j} v_j, \qquad R_{nx} = I_p + \frac{C}{n}.$$

Under Assumption 1, this is the standard local-to-unity setup. Hence the usual functional limit theory for nearly integrated autoregressions implies that the process

$$X_n(r) := n^{-1/2} x_{\lfloor nr \rfloor}, \qquad 0 \le r \le 1,$$

is tight in $D[0,1]$ and in fact converges weakly to the continuous Ornstein-Uhlenbeck-type limit associated with C (see, for example, Phillips and Magdalinos, 2007). By the continuous-mapping theorem,

$$\sup_{0 \le r \le 1} \|X_n(r)\| = O_p(1),$$

which is equivalent to

$$\max_{1 \le t \le n} \|x_t\| = O_p(n^{1/2}).$$

Next, by the IVX recursion (11),

$$\tilde{z}_t = \sum_{j=1}^{t} R_{nz}^{t-j} \Delta x_j, \qquad R_{nz} = I_p + \frac{C_z}{n^\delta}.$$

Using the predictor recursion in (18),

$$\Delta x_t = x_t - x_{t-1} = \frac{C}{n} x_{t-1} + v_t.$$

Hence we may decompose

$$\tilde{z}_t = z_t^{(v)} + r_{t,n}, \qquad z_t^{(v)} := \sum_{j=1}^{t} R_{nz}^{t-j} v_j, \qquad r_{t,n} := \sum_{j=1}^{t} R_{nz}^{t-j} \frac{C}{n} x_{j-1}.$$

Because the eigenvalues of C_z have strictly negative real parts, there exist constants $M, c > 0$ such that

$$\|R_{nz}^k\| \le M \exp\Big(-ck/n^\delta\Big), \qquad k \ge 0,$$

for all sufficiently large n (Magdalinos and Phillips, 2009; Kostakis et al., 2015). Therefore,

$$\max_{1\le t\le n} \|r_{t,n}\| \le \frac{\|C\|}{n} \max_{0\le j\le n-1} \|x_j\| \max_{1\le t\le n} \sum_{j=1}^{t} \|R_{nz}^{t-j}\| \le \frac{M\|C\|}{n} \max_{0\le j\le n-1} \|x_j\| \sum_{k=0}^{n-1} \exp\Big(-ck/n^\delta\Big).$$

The geometric sum on the right-hand side is $O(n^\delta)$, while the first step of the proof established that $\max_{0\le j\le n-1} \|x_j\| = O_p(n^{1/2})$. Consequently,

$$\max_{1\le t\le n} \|r_{t,n}\| = O_p\Big(n^{\delta-1/2}\Big) = o_p\Big(n^{\delta/2}\Big),$$

because $\delta < 1$.

It therefore remains to control the innovation-driven component $z_t^{(v)}$. For this term, the input is the martingale-difference innovation v_t and the recursion is exactly the mildly integrated/IVX filter studied in Magdalinos and Phillips (2009), Kostakis et al. (2015), and Lee (2016). Standard mildly integrated bounds therefore yield tightness of

$$Z_n^{(v)}(r) := n^{-\delta/2} z_{\lfloor nr \rfloor}^{(v)}, \qquad 0 \le r \le 1,$$

again with a continuous normal limit. Consequently,

$$\max_{1\le t\le n} \|z_t^{(v)}\| = O_p(n^{\delta/2}).$$

Combining this with the bound for $r_{t,n}$ yields

$$\max_{1\le t\le n} \|\tilde{z}_t\| \le \max_{1\le t\le n} \|z_t^{(v)}\| + \max_{1\le t\le n} \|r_{t,n}\| = O_p(n^{\delta/2}).$$

For the asymptotic normality of the normalized score, define

$$\xi_{t,n} := n^{-1/2} \bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau), \qquad 1 \le t \le n-1.$$

By Assumption 2(ii), $\{\xi_{t,n}, \mathcal{F}_t\}$ is a martingale-difference sequence. Its conditional covariance sum satisfies

$$\sum_{t=1}^{n-1} \mathbb{E}\Big[\xi_{t,n}\xi_{t,n}' \mid \mathcal{F}_t\Big] = \frac{1}{n}\sum_{t=1}^{n-1} \mathbb{E}\Big[\bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau)\bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau)' \mid \mathcal{F}_t\Big] \xrightarrow{p} \Omega_\tau^{\mathrm{IVX}}$$

by Assumption 2(v). Moreover, Assumption 2(iii) implies the Lyapunov bound

$$\sum_{t=1}^{n-1} \mathbb{E}\Big[\|\xi_{t,n}\|^{2+\eta}\Big] = n^{-(2+\eta)/2}\sum_{t=1}^{n-1} \mathbb{E}\Big[\|\bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau)\|^{2+\eta}\Big] \le n^{-\eta/2}\sup_{n,t} \mathbb{E}\Big[\|\bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau)\|^{2+\eta}\Big] \to 0.$$

Hence the conditional Lindeberg condition holds. The multivariate martingale central limit theorem then gives

$$\sum_{t=1}^{n-1} \xi_{t,n} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1} \bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau) \xrightarrow{d} N\Big(0, \Omega_\tau^{\mathrm{IVX}}\Big),$$

which completes the proof. □

Proof of Lemma 2. For clarity, the proof separates the Jacobian and HAC claims. The Jacobian claim invokes the consistency conclusion of Theorem 1, which is stated later for expository convenience; there is no circularity because the proof of that theorem below is based directly on Assumption 3 and does not use the HAC claim established here.

For the Jacobian, Theorem 1 below implies $\hat{\theta}_\tau^{\mathrm{IVX}} \xrightarrow{p} \theta_\tau$. Since $\mathcal{N}_\tau$ is a neighborhood of θ_τ, this yields

$$\Pr\Big(\hat{\theta}_\tau^{\mathrm{IVX}} \in \mathcal{N}_\tau\Big) \to 1.$$

Assumption 3 then gives the local uniform convergence

$$\sup_{\theta\in\mathcal{N}_\tau} \Big\| S_n \hat{J}_n^{\mathrm{IVX}}(\theta) D_n^{-1} - J_\tau^{\mathrm{IVX}} \Big\| \xrightarrow{p} 0.$$

Evaluating this display at $\theta = \hat{\theta}_\tau^{\mathrm{IVX}}$ on the event $\{\hat{\theta}_\tau^{\mathrm{IVX}} \in \mathcal{N}_\tau\}$ yields

$$\Big\| S_n \hat{J}_\tau^{\mathrm{IVX}} D_n^{-1} - J_\tau^{\mathrm{IVX}} \Big\| \le \sup_{\theta\in\mathcal{N}_\tau} \Big\| S_n \hat{J}_n^{\mathrm{IVX}}(\theta) D_n^{-1} - J_\tau^{\mathrm{IVX}} \Big\| \xrightarrow{p} 0.$$

For the HAC part, first work with the infeasible normalized score at the true parameter,

$$\bar{g}_{t,n} := \bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau) = H_n z_t \psi_\tau(u_{t+1,\tau}),$$

and define the associated sample autocovariances

$$\bar{\Gamma}_{j,n} := \frac{1}{n} \sum_{t=j+1}^{n-1} \bar{g}_{t,n} \bar{g}'_{t-j,n}.$$

The corresponding normalized HAC estimator is

$$\bar{\Omega}_n := \bar{\Gamma}_{0,n} + \sum_{j=1}^{L_n} K\left(\frac{j}{L_n+1}\right) \left(\bar{\Gamma}_{j,n} + \bar{\Gamma}'_{j,n}\right).$$

Define

$$d_{t,n} := \bar{g}_{t,n} \bar{g}'_{t,n} - \mathbb{E}\left[\bar{g}_{t,n} \bar{g}'_{t,n} \mid \mathcal{F}_t\right].$$

Then $\{d_{t,n}, \mathcal{F}_t\}$ is a matrix-valued martingale-difference array. By Assumption 2(iii), its entries have uniformly bounded $(1+\eta/2)$ moments, so the martingale weak law gives

$$\frac{1}{n} \sum_{t=1}^{n-1} d_{t,n} = o_p(1).$$

Combining this with Assumption 2(v),

$$\bar{\Gamma}_{0,n} = \frac{1}{n} \sum_{t=1}^{n-1} \mathbb{E}\left[\bar{g}_{t,n} \bar{g}'_{t,n} \mid \mathcal{F}_t\right] + o_p(1) \xrightarrow{p} \Omega_\tau^{\mathrm{IVX}}.$$

For every fixed $j \geq 1$, $\bar{g}_{t-j,n}$ is measurable with respect to $\mathcal{F}_t$, and Assumption 2(ii) implies

$$\mathbb{E}\left[\bar{g}_{t,n} \bar{g}'_{t-j,n} \mid \mathcal{F}_t\right] = \mathbb{E}[\bar{g}_{t,n} \mid \mathcal{F}_t]\, \bar{g}'_{t-j,n} = 0.$$

Hence each lag-j covariance average is centered. More precisely, for every $j \geq 1$ and every matrix entry, the array $\{\bar{g}_{t,n} \bar{g}'_{t-j,n}, \mathcal{F}_t\}$ is a martingale difference. A standard martingale variance calculation, using orthogonality of the increments together with Assumption 2(iii), yields

$$\mathrm{Var}\Big(\mathrm{vec}(\bar{\Gamma}_{j,n})\Big) = O\left(\frac{1}{n}\right)$$

uniformly over $1 \leq j \leq L_n$. Equivalently,

$$\bar{\Gamma}_{j,n} = O_p\left(\frac{1}{\sqrt{n}}\right)$$

uniformly over $1 \leq j \leq L_n$. Since $K(\cdot)$ is bounded,

$$\sum_{j=1}^{L_n} \left\| K\left(\frac{j}{L_n+1}\right)\left(\bar{\Gamma}_{j,n} + \bar{\Gamma}'_{j,n}\right)\right\| = O_p\left(\frac{L_n}{\sqrt{n}}\right) = o_p(1),$$

and therefore

$$\bar{\Omega}_n \xrightarrow{p} \Omega_\tau^{\mathrm{IVX}}.$$

Finally, the feasible HAC estimator is asymptotically equivalent to the infeasible one. Since

$$\psi_\tau(u) = \tau u \mathbf{1}(u \geq 0) + (1-\tau) u \mathbf{1}(u < 0),$$

the expectile score is globally Lipschitz with constant $L_\psi := \max\{\tau, 1-\tau\}$. Therefore

$$\left\| H_n \hat{g}_t^{\mathrm{IVX}} - \bar{g}_{t,n} \right\| \leq L_\psi \|H_n z_t\| \left| r_t' \left(\hat{\theta}_\tau^{\mathrm{IVX}} - \theta_\tau\right)\right|.$$

By Lemma 1, $\max_t \|H_n z_t\| = O_p(1)$ and $\max_t \|x_t\| = O_p(n^{1/2})$. By Theorem 1,

$$|\hat{\alpha}_\tau^{\mathrm{IVX}} - \alpha_\tau| = O_p(n^{-1/2}), \qquad \|\hat{\beta}_\tau^{\mathrm{IVX}} - \beta_\tau\| = O_p(n^{-1}).$$

Consequently,

$$\max_{1\leq t\leq n} \left| r_t' \left(\hat{\theta}_\tau^{\mathrm{IVX}} - \theta_\tau\right)\right| \leq |\hat{\alpha}_\tau^{\mathrm{IVX}} - \alpha_\tau| + \max_{1\leq t\leq n} \|x_t\| \|\hat{\beta}_\tau^{\mathrm{IVX}} - \beta_\tau\| = O_p(n^{-1/2}),$$

and therefore

$$\max_{1\leq t\leq n} \left\| H_n \hat{g}_t^{\mathrm{IVX}} - \bar{g}_{t,n} \right\| = O_p(n^{-1/2}).$$

Let

$$\Delta_n := \max_{1\leq t\leq n} \left\| H_n \hat{g}_t^{\mathrm{IVX}} - \bar{g}_{t,n} \right\|, \qquad \tilde{\Gamma}_{j,n} := \frac{1}{n}\sum_{t=j+1}^{n-1} H_n \hat{g}_t^{\mathrm{IVX}} \left(H_n \hat{g}_{t-j}^{\mathrm{IVX}}\right)'.$$

Then, for each $0 \leq j \leq L_n$,

$$\left\| \tilde{\Gamma}_{j,n} - \bar{\Gamma}_{j,n} \right\| \leq \Delta_n \frac{1}{n}\sum_{t=j+1}^{n-1} \left\| H_n \hat{g}_{t-j}^{\mathrm{IVX}} \right\| + \Delta_n \frac{1}{n}\sum_{t=j+1}^{n-1} \|\bar{g}_{t,n}\|.$$

By Assumption 2(iii),

$$\frac{1}{n}\sum_{t=1}^{n-1} \|\bar{g}_{t,n}\| = O_p(1).$$

Also,

$$\frac{1}{n}\sum_{t=1}^{n-1}\left\|H_n\hat{g}_t^{\mathrm{IVX}}\right\| \leq \frac{1}{n}\sum_{t=1}^{n-1}\|\bar{g}_{t,n}\| + \Delta_n = O_p(1).$$

Since these bounds are uniform in j, we obtain

$$\sup_{0\leq j\leq L_n}\left\|\tilde{\Gamma}_{j,n} - \bar{\Gamma}_{j,n}\right\| = O_p(n^{-1/2}).$$

Therefore

$$H_n\hat{\Omega}_\tau^{\mathrm{IVX}}H_n' - \bar{\Omega}_n = O_p\left(\frac{L_n}{\sqrt{n}}\right) = o_p(1),$$

and therefore

$$H_n\hat{\Omega}_\tau^{\mathrm{IVX}}H_n' \xrightarrow{p} \Omega_\tau^{\mathrm{IVX}}.$$

This completes the proof. $\square$

Proof of Theorem 1. For $h \in \mathbb{R}^{p+1}$, define the locally rescaled moment map

$$\Psi_n(h) := S_n g_n^{\mathrm{IVX}}\left(\theta_\tau + D_n^{-1}h\right).$$

For each t and each fixed h, the map

$$s \mapsto \psi_\tau\left(y_{t+1} - r_t'\left(\theta_\tau + sD_n^{-1}h\right)\right)$$

is absolutely continuous, and Assumption 2(iv) rules out the event that the residual equals the kink point with positive probability. Hence a component-wise mean-value expansion yields

$$\Psi_n(h) = \Psi_n(0) - \mathcal{J}_n(h)h, \qquad \mathcal{J}_n(h) := \int_0^1 S_n\hat{J}_n^{\mathrm{IVX}}\left(\theta_\tau + sD_n^{-1}h\right)D_n^{-1}ds.$$

Here $\mathcal{J}_n(h)$ is understood as the average of the per-observation mean-value Jacobians; the integral representation is a convenient shorthand for this local linearization. By Lemma 1,

$$\Psi_n(0) = S_n g_n^{\mathrm{IVX}}(\theta_\tau) = \frac{n}{n-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n-1}\bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau) = O_p(1).$$

Moreover, for every fixed $M > 0$,

$$\sup_{\|h\| \le M} \left\| \mathcal{J}_n(h) - J_\tau^{\mathrm{IVX}} \right\| \xrightarrow{p} 0,$$

because $D_n^{-1} h \to 0$ uniformly on bounded sets and Assumption 3 gives local uniform convergence of the normalized Jacobian on a neighborhood of θ_τ. In particular, for every fixed M, we have $\theta_\tau + s D_n^{-1} h \in \mathcal{N}_\tau$ for all $s \in [0,1]$ and $\|h\| \le M$ with probability approaching one.

We first establish existence of a root with $\|h\| = O_p(1)$. Fix $\varepsilon > 0$. Since $\Psi_n(0) = O_p(1)$ and J_τ^{IVX} is nonsingular, we can choose $M < \infty$ such that

$$\Pr\left(\left\| \left(J_\tau^{\mathrm{IVX}} \right)^{-1} \Psi_n(0) \right\| \le \frac{M}{2} \right) \ge 1 - \varepsilon$$

for all large n. Also, by the previous expression,

$$\Pr\left(\sup_{\|h\| \le M} \left\| \left(J_\tau^{\mathrm{IVX}} \right)^{-1} \left[\mathcal{J}_n(h) - J_\tau^{\mathrm{IVX}} \right] h \right\| \le \frac{M}{2} \right) \ge 1 - \varepsilon$$

for all large n. On the intersection of these two events, the continuous map

$$T_n(h) := \left(J_\tau^{\mathrm{IVX}} \right)^{-1} \Psi_n(0) + \left[I - \left(J_\tau^{\mathrm{IVX}} \right)^{-1} \mathcal{J}_n(h) \right] h$$

maps the closed ball $B_M := \{h : \|h\| \le M\}$ into itself. By Brouwer's fixed-point theorem, there exists $h_n \in B_M$ such that $T_n(h_n) = h_n$, which is equivalent to

$$\Psi_n(h_n) = 0.$$

Define

$$\hat{\theta}_\tau^{\mathrm{IVX}} := \theta_\tau + D_n^{-1} h_n.$$

Then $\hat{\theta}_\tau^{\mathrm{IVX}}$ solves (14) and

$$D_n \left(\hat{\theta}_\tau^{\mathrm{IVX}} - \theta_\tau \right) = h_n = O_p(1).$$

Since $D_n^{-1} \to 0$ component-wise, it follows that

$$\hat{\theta}_\tau^{\mathrm{IVX}} \xrightarrow{p} \theta_\tau.$$

To obtain the asymptotic linear representation, let $\Delta_n := D_n(\hat{\theta}_\tau^{\mathrm{IVX}} - \theta_\tau) = O_p(1)$. Using the same mean-value expansion evaluated at $h = \Delta_n$ and the identity $\Psi_n(\Delta_n) = 0$,

$$0 = \Psi_n(0) - \mathcal{J}_n(\Delta_n)\Delta_n, \qquad \Delta_n = \mathcal{J}_n(\Delta_n)^{-1}\Psi_n(0).$$

Because $\Delta_n = O_p(1)$, for every $\varepsilon > 0$ we can choose $M < \infty$ such that $\Pr(\|\Delta_n\| > M) < \varepsilon$ for all large n. On the event $\{\|\Delta_n\| \le M\}$, the uniform convergence of $\mathcal{J}_n(h)$ on bounded sets applies, and therefore

$$\mathcal{J}_n(\Delta_n) \xrightarrow{p} J_\tau^{\mathrm{IVX}},$$

Since J_τ^{IVX} is nonsingular, this also implies that $\mathcal{J}_n(\Delta_n)$ is invertible with probability approaching one and

$$\mathcal{J}_n(\Delta_n)^{-1} = \left(J_\tau^{\mathrm{IVX}}\right)^{-1} + o_p(1).$$

Substituting this into the previous equation gives

$$\Delta_n = \left(J_\tau^{\mathrm{IVX}}\right)^{-1} \Psi_n(0) + o_p(1).$$

Finally,

$$\Psi_n(0) = \frac{n}{n-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n-1} \bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1} \bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau) + o_p(1),$$

because the score sum is $O_p(1)$ by Lemma 1. Therefore

$$D_n\left(\hat{\theta}_\tau^{\mathrm{IVX}} - \theta_\tau\right) = \left(J_\tau^{\mathrm{IVX}}\right)^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1} \bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau) + o_p(1).$$

Since $D_n = \mathrm{diag}(n^{1/2}, nI_p)$, the component-wise rates follow

$$\hat{\alpha}_\tau^{\mathrm{IVX}} - \alpha_\tau = O_p(n^{-1/2}), \qquad \hat{\beta}_\tau^{\mathrm{IVX}} - \beta_\tau = O_p(n^{-1}).$$

This completes the proof. $\square$

Proof of Theorem 2. The first claim follows directly from Theorem 1 and Lemma 1. Indeed, Theorem 1 gives the asymptotic linear representation

$$D_n\Big(\hat{\theta}_\tau^{\mathrm{IVX}} - \theta_\tau\Big) = \Big(J_\tau^{\mathrm{IVX}}\Big)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n-1} \bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau) + o_p(1),$$

while Lemma 1 implies

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n-1} \bar{g}_{t,n}^{\mathrm{IVX}}(\theta_\tau) \xrightarrow{d} N\Big(0, \Omega_\tau^{\mathrm{IVX}}\Big).$$

Therefore, by Slutsky's theorem,

$$D_n\Big(\hat{\theta}_\tau^{\mathrm{IVX}} - \theta_\tau\Big) \xrightarrow{d} N\bigg(0, \Big(J_\tau^{\mathrm{IVX}}\Big)^{-1} \Omega_\tau^{\mathrm{IVX}} \Big(J_\tau^{\mathrm{IVX}}\Big)^{-1\prime}\bigg) = N\Big(0, V_\tau^{\mathrm{IVX}}\Big).$$

For the covariance estimator, note first that

$$D_n \hat{V}_\tau^{\mathrm{IVX}} D_n = \Big(D_n \hat{J}_\tau^{\mathrm{IVX}\,-1} S_n^{-1}\Big)\Big(H_n \hat{\Omega}_\tau^{\mathrm{IVX}} H_n'\Big)\Big(D_n \hat{J}_\tau^{\mathrm{IVX}\,-1} S_n^{-1}\Big)'.$$

Since

$$\Big(S_n \hat{J}_\tau^{\mathrm{IVX}} D_n^{-1}\Big)^{-1} = D_n \hat{J}_\tau^{\mathrm{IVX}\,-1} S_n^{-1},$$

Lemma 2 implies

$$D_n \hat{J}_\tau^{\mathrm{IVX}\,-1} S_n^{-1} \xrightarrow{p} \Big(J_\tau^{\mathrm{IVX}}\Big)^{-1}$$

and

$$H_n \hat{\Omega}_\tau^{\mathrm{IVX}} H_n' \xrightarrow{p} \Omega_\tau^{\mathrm{IVX}}.$$

Applying the continuous mapping theorem to these probability limits,

$$D_n \hat{V}_\tau^{\mathrm{IVX}} D_n \xrightarrow{p} \Big(J_\tau^{\mathrm{IVX}}\Big)^{-1} \Omega_\tau^{\mathrm{IVX}} \Big(J_\tau^{\mathrm{IVX}}\Big)^{-1\prime} = V_\tau^{\mathrm{IVX}}.$$

This completes the proof. □

Proof of Theorem 3. Under $\mathbb{H}_0\colon R\theta_\tau = q$, define

$$Z_n := D_n\Big(\hat{\theta}_\tau^{\mathrm{IVX}} - \theta_\tau\Big), \qquad \Sigma_n := D_n \hat{V}_\tau^{\mathrm{IVX}} D_n, \qquad M_n := R D_n^{-1}.$$

Then

$$R\hat{\theta}_\tau^{\mathrm{IVX}} - q = R\Big(\hat{\theta}_\tau^{\mathrm{IVX}} - \theta_\tau\Big) = M_n Z_n,$$

and therefore the Wald statistic can be written as

$$W_{n,\tau}^{\mathrm{IVX}} = Z_n' M_n' \left(M_n \Sigma_n M_n'\right)^{-1} M_n Z_n.$$

By Theorem 2,

$$Z_n \xrightarrow{d} N\Big(0, V_\tau^{\mathrm{IVX}}\Big), \qquad \Sigma_n \xrightarrow{p} V_\tau^{\mathrm{IVX}}.$$

Write $V := V_\tau^{\mathrm{IVX}}$ for brevity. Since $\Omega_\tau^{\mathrm{IVX}}$ is positive definite and J_τ^{IVX} is nonsingular, V is positive definite. Define

$$U_n := V^{-1/2} Z_n.$$

Then $U_n \xrightarrow{d} N(0, I_{p+1})$. Also define the deterministic rank-m projector

$$P_n := V^{1/2} M_n' \left(M_n V M_n'\right)^{-1} M_n V^{1/2}.$$

Because R has full row rank and D_n is diagonal with strictly positive entries, M_n has full row rank for every n, so $M_n V M_n'$ is invertible. Moreover,

$$P_n' = P_n, \qquad P_n^2 = P_n, \qquad \mathrm{rank}(P_n) = m.$$

We first show that the feasible Wald statistic is asymptotically equivalent to the quadratic form $U_n' P_n U_n$. Let

$$A_n := \left(M_n V M_n'\right)^{-1/2} M_n V^{1/2}, \qquad B_n := V^{-1/2} \Sigma_n V^{-1/2},$$

so that

$$A_n A_n' = I_m, \qquad P_n = A_n' A_n, \qquad B_n \xrightarrow{p} I_{p+1}$$

by Theorem 2. Moreover,

$$W_{n,\tau}^{\mathrm{IVX}} = U_n' A_n' \left(A_n B_n A_n'\right)^{-1} A_n U_n.$$

Since $A_n A_n' = I_m$,

$$A_n B_n A_n' = I_m + A_n(B_n - I_{p+1})A_n',$$

and therefore

$$\|A_n B_n A_n' - I_m\| \le \|B_n - I_{p+1}\| \xrightarrow{p} 0.$$

Hence

$$(A_n B_n A_n')^{-1} = I_m + o_p(1).$$

It follows that

$$W_{n,\tau}^{\mathrm{IVX}} = U_n' P_n U_n + U_n' A_n' o_p(1) A_n U_n.$$

Because

$$\|A_n\|^2 = \lambda_{\max}(A_n' A_n) = \lambda_{\max}(P_n) = 1$$

and $U_n = O_p(1)$, the remainder term satisfies

$$|U_n' A_n' o_p(1) A_n U_n| \le o_p(1)\|A_n U_n\|^2 \le o_p(1)\|U_n\|^2 = o_p(1).$$

Therefore

$$W_{n,\tau}^{\mathrm{IVX}} = U_n' P_n U_n + o_p(1).$$

It remains to show that $U_n' P_n U_n \xrightarrow{d} \chi_m^2$. Because $M_n = R D_n^{-1}$ combines the $n^{-1/2}$ normalization for the intercept with the n^{-1} normalization for the slopes, the projector P_n need not converge when R mixes intercept and slope restrictions, so we proceed by subsequences. The set

$$\mathcal{P}_m := \left\{ P \in \mathbb{R}^{(p+1)\times(p+1)} : P' = P,\ P^2 = P,\ \mathrm{rank}(P) = m \right\}$$

is compact: every $P \in \mathcal{P}_m$ is a projector and hence bounded in norm by one, while closedness follows because limits preserve symmetry and idempotence, and $\mathrm{tr}(P_n) = m$ for all n implies $\mathrm{tr}(P) = m$ at the limit, so $\mathrm{rank}(P) = m$. Hence every subsequence of $\{P_n\}$ contains a further subsequence, still denoted by $\{P_n\}$, such that

$$P_n \to P \in \mathcal{P}_m.$$

Along the same subsequence,

$$(U_n, P_n) \xrightarrow{d} (U, P),$$

where $U \sim N(0, I_{p+1})$ and P is deterministic. By the continuous-mapping theorem,

$$U_n' P_n U_n \xrightarrow{d} U' P U.$$

Since $P \in \mathcal{P}_m$, there exists an orthogonal matrix Q such that

$$P = Q' \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} Q.$$

Because U follows standard normal and hence is rotationally invariant,

$$U' P U = (QU)' \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} (QU) \stackrel{d}{=} \sum_{j=1}^{m} U_j^2 \sim \chi_m^2.$$

Therefore every subsequence of $U_n' P_n U_n$ has a further subsequence converging to χ_m^2, which implies

$$U_n' P_n U_n \xrightarrow{d} \chi_m^2.$$

Combining this with the asymptotic equivalence above yields

$$W_{n,\tau}^{\mathrm{IVX}} \xrightarrow{d} \chi_m^2.$$

The size statement follows immediately from convergence to the χ_m^2 critical values. The corresponding confidence region is obtained by inverting the Wald test, so its asymptotic coverage is $1 - \alpha$. This completes the proof. $\square$

B Additional Simulation Results

Figure A1: Comparison of size-adjusted power: Univariate model with $t(5)$ disturbance

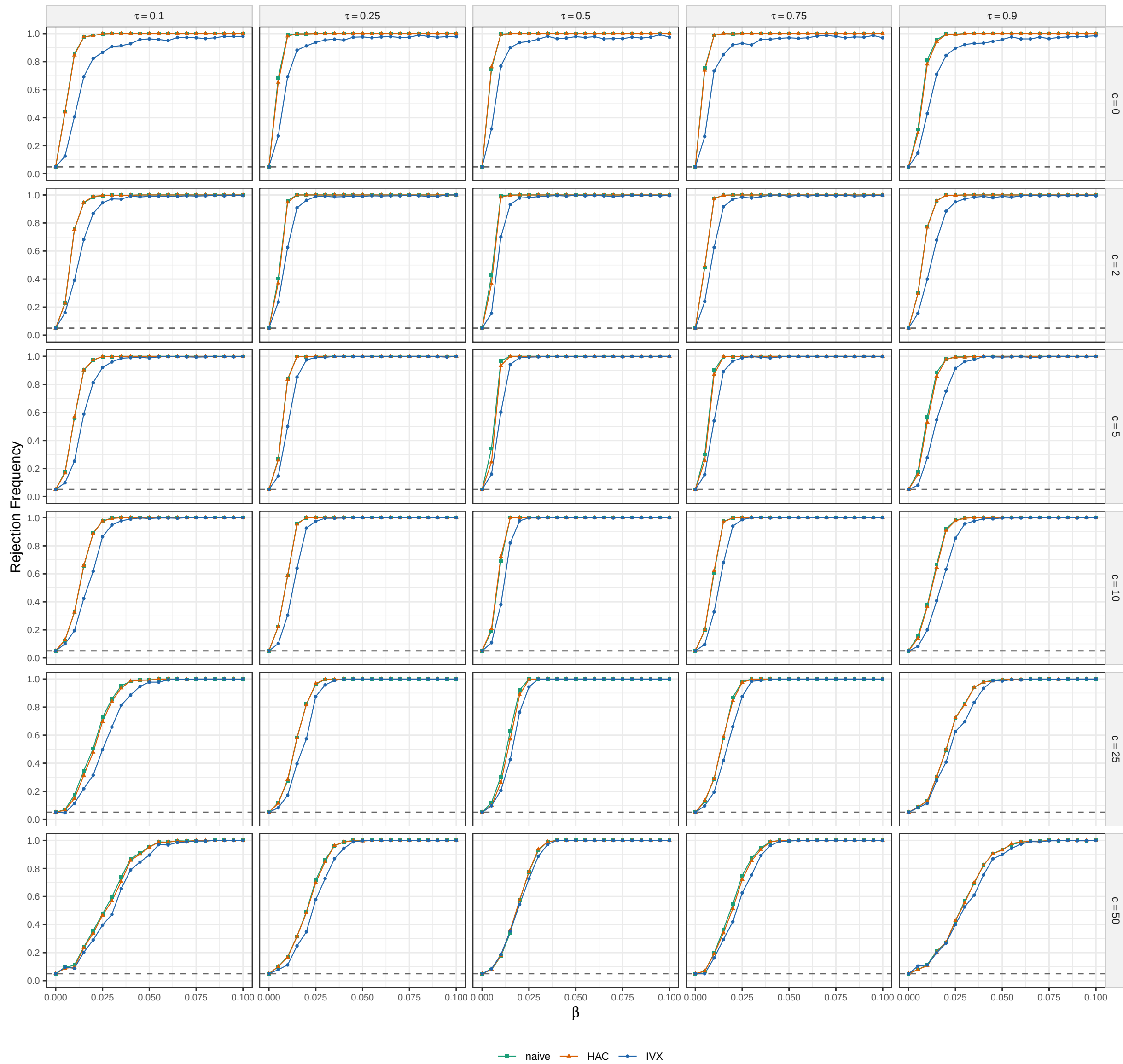


Figure A2: Comparison of size-adjusted power: Univariate model with GARCH disturbance

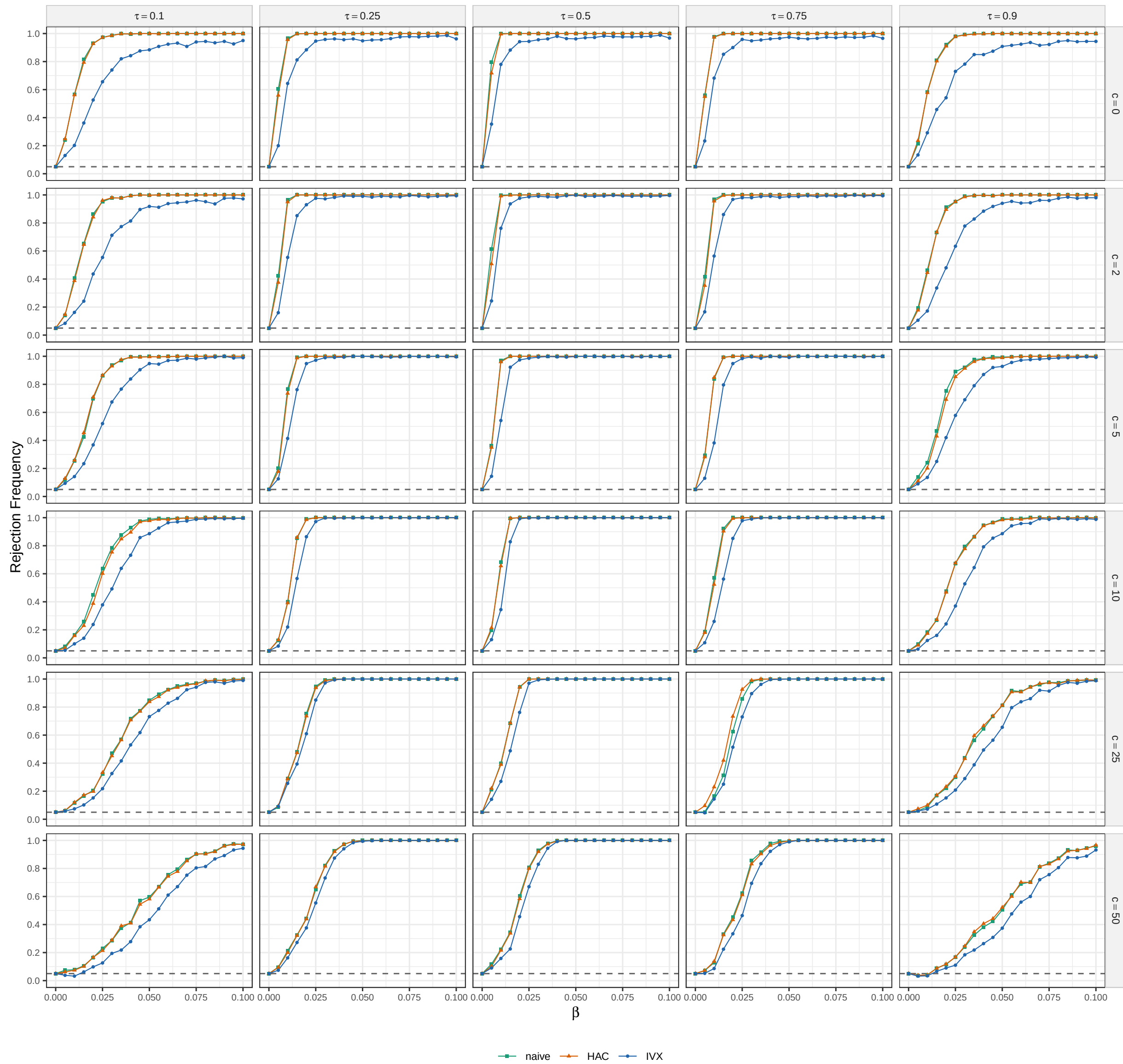


Table A1: Empirical rejection frequencies under the null: Univariate model under $t(5)$ and GARCH disturbance

		$\tau = 0.1$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.9$		
T	c	naive	HAC	IVX	naive	HAC	IVX	naive	HAC	IVX	naive	HAC	IVX	naive	HAC	IVX
Panel A. Student's $t(5)$																
250	0	0.1860	0.2015	0.0590	0.2340	0.2645	0.0545	0.2825	0.3000	0.0575	0.2615	0.2830	0.0550	0.2080	0.2270	0.0680
	2	0.1375	0.1460	0.0695	0.1635	0.1845	0.0555	0.1740	0.1965	0.0605	0.1525	0.1645	0.0595	0.1295	0.1355	0.0570
	5	0.1055	0.1150	0.0660	0.1040	0.1250	0.0480	0.1095	0.1200	0.0480	0.1135	0.1250	0.0500	0.1055	0.1085	0.0585
	10	0.0750	0.0790	0.0610	0.0785	0.0885	0.0535	0.0925	0.1020	0.0480	0.0865	0.0980	0.0660	0.0850	0.0890	0.0715
	25	0.0690	0.0745	0.0615	0.0650	0.0715	0.0425	0.0720	0.0820	0.0575	0.0710	0.0780	0.0565	0.0675	0.0755	0.0705
	50	0.0650	0.0785	0.0720	0.0640	0.0805	0.0570	0.0630	0.0785	0.0595	0.0575	0.0670	0.0530	0.0695	0.0850	0.0590
500	0	0.1825	0.1935	0.0665	0.2470	0.2655	0.0500	0.2840	0.2970	0.0540	0.2630	0.2730	0.0530	0.1810	0.2000	0.0555
	2	0.1340	0.1420	0.0545	0.1525	0.1595	0.0565	0.1715	0.1830	0.0485	0.1505	0.1635	0.0490	0.1230	0.1325	0.0560
	5	0.0875	0.0915	0.0520	0.1120	0.1230	0.0550	0.1070	0.1210	0.0530	0.1165	0.1190	0.0465	0.0905	0.0990	0.0555
	10	0.0760	0.0850	0.0530	0.0920	0.0975	0.0460	0.0805	0.0920	0.0420	0.0745	0.0795	0.0455	0.0695	0.0765	0.0585
	25	0.0585	0.0680	0.0565	0.0600	0.0715	0.0520	0.0575	0.0620	0.0390	0.0560	0.0585	0.0495	0.0655	0.0805	0.0645
	50	0.0680	0.0770	0.0615	0.0550	0.0620	0.0535	0.0655	0.0710	0.0535	0.0405	0.0490	0.0425	0.0570	0.0625	0.0550
1000	0	0.1885	0.1970	0.0460	0.2430	0.2520	0.0580	0.2840	0.2965	0.0495	0.2260	0.2280	0.0495	0.1960	0.2015	0.0505
	2	0.1170	0.1215	0.0545	0.1570	0.1680	0.0555	0.1850	0.1855	0.0560	0.1615	0.1700	0.0565	0.1180	0.1250	0.0570
	5	0.0840	0.0850	0.0485	0.1035	0.1045	0.0540	0.1255	0.1340	0.0640	0.1085	0.1120	0.0500	0.0845	0.0930	0.0490
	10	0.0740	0.0820	0.0520	0.0855	0.0905	0.0425	0.0810	0.0895	0.0455	0.0845	0.0900	0.0435	0.0710	0.0765	0.0570
	25	0.0625	0.0675	0.0565	0.0715	0.0740	0.0560	0.0605	0.0625	0.0530	0.0590	0.0620	0.0555	0.0745	0.0770	0.0590
	50	0.0540	0.0600	0.0510	0.0525	0.0600	0.0535	0.0570	0.0575	0.0475	0.0475	0.0520	0.0505	0.0570	0.0660	0.0525
Panel B. GARCH																
250	0	0.1910	0.2215	0.0725	0.2615	0.2775	0.0630	0.2785	0.3020	0.0530	0.2690	0.2915	0.0530	0.1885	0.2070	0.0565
	2	0.1145	0.1330	0.0540	0.1500	0.1625	0.0585	0.1685	0.1755	0.0455	0.1520	0.1670	0.0565	0.1120	0.1260	0.0580
	5	0.1070	0.1140	0.0640	0.1080	0.1220	0.0540	0.1090	0.1305	0.0505	0.1100	0.1170	0.0595	0.1010	0.1100	0.0600
	10	0.0895	0.0930	0.0700	0.0915	0.1095	0.0595	0.0910	0.0995	0.0535	0.0880	0.1005	0.0530	0.0805	0.0930	0.0600
	25	0.0685	0.0825	0.0645	0.0670	0.0865	0.0530	0.0705	0.0845	0.0505	0.0650	0.0765	0.0540	0.0715	0.0815	0.0685
	50	0.0695	0.0800	0.0645	0.0615	0.0770	0.0545	0.0660	0.0745	0.0600	0.0615	0.0700	0.0540	0.0840	0.0885	0.0795
500	0	0.1675	0.1805	0.0615	0.2600	0.2745	0.0510	0.2900	0.3005	0.0620	0.2430	0.2515	0.0530	0.1675	0.1800	0.0545
	2	0.1265	0.1320	0.0650	0.1570	0.1675	0.0535	0.1615	0.1750	0.0525	0.1490	0.1615	0.0595	0.1285	0.1370	0.0600
	5	0.0905	0.1000	0.0630	0.1180	0.1230	0.0650	0.0990	0.1125	0.0480	0.0985	0.1090	0.0480	0.0850	0.0960	0.0545
	10	0.0755	0.0810	0.0690	0.0740	0.0825	0.0530	0.0810	0.0895	0.0490	0.0875	0.0930	0.0420	0.0795	0.0895	0.0580
	25	0.0690	0.0795	0.0540	0.0655	0.0770	0.0550	0.0730	0.0830	0.0600	0.0640	0.0690	0.0560	0.0595	0.0660	0.0480
	50	0.0625	0.0650	0.0590	0.0515	0.0595	0.0470	0.0580	0.0660	0.0545	0.0630	0.0665	0.0495	0.0610	0.0710	0.0555
1000	0	0.1825	0.1880	0.0675	0.2505	0.2520	0.0590	0.2760	0.2860	0.0570	0.2485	0.2615	0.0555	0.1775	0.1895	0.0620
	2	0.1075	0.1170	0.0495	0.1590	0.1605	0.0560	0.1785	0.1700	0.0595	0.1500	0.1615	0.0430	0.1215	0.1275	0.0555
	5	0.0845	0.0955	0.0525	0.0995	0.1025	0.0505	0.1160	0.1230	0.0570	0.1045	0.1125	0.0515	0.0780	0.0885	0.0565
	10	0.0660	0.0700	0.0545	0.0850	0.0940	0.0455	0.0795	0.0845	0.0450	0.0855	0.0930	0.0545	0.0725	0.0755	0.0530
	25	0.0610	0.0675	0.0570	0.0590	0.0650	0.0445	0.0745	0.0760	0.0495	0.0665	0.0755	0.0485	0.0585	0.0645	0.0560
	50	0.0490	0.0520	0.0480	0.0565	0.0620	0.0465	0.0610	0.0625	0.0495	0.0525	0.0525	0.0430	0.0485	0.0545	0.0490

C Additional Empirical Results

Table A2: Unit root tests for predictive regressors: Monthly and quarterly data

	Full sample: 1927–2024			Subsample: 1952–2024		
Predictor	$\hat{\theta}_1$	ADF	PP	$\hat{\theta}_1$	ADF	PP
Panel A. Monthly						
Dividend-earnings ratio	0.9920	-6.0666***	-4.5561***	0.9873	-6.5159***	-4.8050***
Long-term yield	0.9961	-1.6609	-1.5093	0.9947	-1.7261	-1.6015
Dividend yield	0.9949	-1.7784	-1.6945	0.9950	-1.5005	-1.5680
Dividend-price ratio	0.9948	-1.7828	-1.7678	0.9949	-1.6228	-1.6136
Treasury bill rate	0.9932	-2.8155*	-2.3359	0.9912	-2.8214*	-2.3784
Earnings-price ratio	0.9881	-3.4606***	-3.6757***	0.9896	-3.0577**	-3.2388**
Book-to-market ratio	0.9881	-3.1688**	-2.5933*	0.9953	-1.4662	-1.5026
Default yield spread	0.9750	-4.8472***	-3.9448***	0.9687	-5.1392***	-4.5283***
Net equity issuance	0.9816	-3.8380***	-4.0925***	0.9819	-3.2048**	-3.5221***
Term spread	0.9644	-5.1757***	-4.7310***	0.9619	-4.6510***	-4.3199***
Inflation	0.4828	-15.0202***	-21.8220***	0.5451	-12.6658***	-16.5118***
Panel B. Quarterly						
Dividend-earnings ratio	0.9323	-6.6595***	-4.6333***	0.8975	-7.4871***	-4.9525***
Long-term yield	0.9862	-1.5314	-1.5958	0.9814	-1.6117	-1.6721
Dividend yield	0.9797	-1.7704	-1.7473	0.9823	-1.7546	-1.6266
Dividend-price ratio	0.9789	-1.8074	-1.9007	0.9812	-1.8052	-1.6748
Treasury bill rate	0.9636	-2.4501	-2.5624	0.9536	-2.4261	-2.5716
Earnings-price ratio	0.9408	-4.2330***	-3.7001***	0.9447	-4.0794***	-3.2531**
Book-to-market ratio	0.9497	-2.4086	-2.9181**	0.9833	-1.7105	-1.5919
Default yield spread	0.8983	-3.4851***	-4.3061***	0.8731	-4.7949***	-4.4478***
Net equity issuance	0.9358	-4.3001***	-4.3982***	0.9374	-3.9275***	-3.7120***
Term spread	0.8656	-5.0503***	-5.3035***	0.8561	-4.5157***	-4.7397***
Inflation	0.4924	-8.6985***	-12.2423***	0.4266	-7.6920***	-11.6662***
Consumption-wealth ratio	–	–	–	0.8508	-3.6148***	-4.4606***

Notes. This table presents the results of unit root tests for the following monthly (panel A) and quarterly (panel B) financial and macroeconomic variables$\hat{\theta}_1$ denotes the least squares estimates of the slope coefficient from the first-order autoregression $x_t = \theta_0 + \theta_1 x_{t-1} + e_t$. ADF and PP denote the augmented Dickey-Fuller (Said and Dickey, 1984) and Phillips-Perron (Phillips and Perron, 1988) unit root test statistics for the null of the presence of a unit root. *, **, and *** indicate rejection at the 10%, 5%, and 1% levels, respectively.