Unified Inference for Predictive Mean and Quantile Regressions via Empirical Likelihood*

Zongwu Cai† Yifeng Chen‡ Seok Young Hong§ Daniel Tsvetanov¶

January 27, 2026

Abstract

We develop an empirical likelihood framework for testing return predictability in the conditional mean and conditional quantiles. A unified chi-square limit theory is established across a broad spectrum of predictor persistence, including stationary, mildly integrated, nearly integrated, unit-root, and mildly explosive cases. We provide two complementary approaches to handle the unknown intercept: (i) a sample-splitting approach under relaxed regularity conditions and (ii) a new two-stage method that improves efficiency and accommodates quantile inference, where sample-splitting is infeasible. We examine the finite-sample bias of the two-stage method, and propose a bias-correction scheme and gradually saturated weights that improve performance under high persistence. Simulation evidence demonstrates that our tests exhibit competitive size and power across persistence classes, with notable gains in quantile predictability. An empirical application to the U.S. stock market shows modest evidence of mean predictability, whereas quantile-based inference reveals stronger and economically relevant predictability in the tails of the return distribution.

JEL codes: C12, C32, C51, C52.

Keywords: Predictive Mean Regression; Predictive Quantile Regression; Empirical Likelihood; Bartlett Bias Correction.

*We appreciate helpful comments and suggestions from Matias Cattaneo, Yoosoon Chang, Gregory Cox, Ekaterina Kazak, Bonsoo Koo, Ji Hyung Lee, Sokbae Simon Lee, Merrick Li, Ming Li, Oliver Linton, Xun Lu, Ruixuan Liu, Kaiji Motegi, Ariel Neufeld, Guangming Pan, Liang Peng, Patrick Pun, Tatevik Sekhposyan, Myunghwan Matt Seo, Zhentao Shi, Sorawoot Srisuma, Wenjie Wang, Yoon-Jae Whang, and seminar participants at the Chinese University of Hong Kong, Kobe University, and the National University of Singapore. We are also grateful to the participants of the 2025 Midwest Econometrics Group conference at the University of Illinois Urbana-Champaign, the 2025 Econometric Society World Congress in Seoul, and the Stochastic Dominance and Quantile-based Methods in Financial Econometrics Workshop at the University of Cambridge in honor of Oliver Linton and Yoon-Jae Whang. Seok Young Hong gratefully acknowledges financial support from Nanyang Technological University, Singapore (SUG grant 023698-00001; CoHASS IGS grant 025604-00001).

†University of Kansas, Lawrence, KS 66045, United States of America ; caiz@ku.edu.

‡Nanyang Technological University, Singapore 639798, Singapore ; yifeng002@e.ntu.edu.sg.

§Nanyang Technological University, Singapore 639798, Singapore ; seokyoung.hong@ntu.edu.sg.

¶Norwich Business School, University of East Anglia, Norwich NR4 7TJ, UK; d.tsvetanov@uea.ac.uk.

1 Introduction

The predictability of asset returns has long been a central topic in financial economics, as it bears directly on foundational questions of market efficiency, asset pricing, and portfolio choice. At the core of these issues lies the behavior of the conditional distribution of financial returns, encompassing not only the first moment but also higher-order characteristics such as tail risk and asymmetry. Understanding how these distributional features relate to available information provides a powerful lens through which key empirical puzzles in financial economics are explored, such as the equity premium puzzle, excess volatility, and time-variation in expected returns.

While much of the literature has focused on the conditional mean of asset returns and its predictability (see Pesaran and Timmermann, 1995; Dangl and Halling, 2012; Gu et al., 2020), growing attention has been directed toward approaches that capture heterogeneity across the conditional distribution, recognizing that predictive relationships may vary at extreme quantiles where downside risk is important. Because mean regressions summarize average predictability across all quantiles, they can obscure such variation and may not adequately capture predictability in specific regions of the distribution, as discussed by Gonzalo and Pitarakis (2012). This has motivated the use of quantile regressions, which provide a more complete view of return predictability (see, among others, Koenker, 2005; Xiao, 2009; Lee, 2016; Fan and Lee, 2019; Cai et al., 2023).

It is essential to recognize, however, that predictive regressions, whether focused on the mean or other distributional features, face significant econometric challenges. One of the most fundamental challenges is the dynamic nature of predictor variables, which may range from strongly stationary to unit-root and even mildly explosive behavior. Greater persistence (i.e., movement toward nonstationarity) introduces severe complications for standard estimation and inference methods, including biased coefficients and nonstandard limiting distributions; see, for example, Phillips (2015).

In practice, the temporal properties of predictors are typically unknown *a priori*, making it difficult to select a method suited to the true data-generating process. This uncertainty highlights the importance of developing unified inference procedures that remain valid across different persistence regimes. Several approaches have been proposed to tackle this challenge, which can broadly be divided into three strands of research. The first method includes the Bonferroni t-test by Cavanagh et al. (2009) and the Bonferroni Q-test by Campbell and Yogo (2006) for mean predictability, along with the quantile extension developed by Maynard et al. (2024). Despite their innovative treatment of the problem, Bonferroni methods face some limitations, including nonstandard asymptotics and joint normality assumptions on the error terms.

The second strand builds on the IVX methodology introduced by Magdalinos and Phillips (2009). This seminal approach involves filtering the predictor to construct an instrumental

variable with a controlled degree of persistence, which is then used in the predictive regression of interest. A key advantage of this method is that the resulting test statistic follows a standard chi-squared limiting distribution, regardless of the persistence characteristics of the original predictor. Several notable contributions have built upon this framework. For example, in the context of mean predictability, Kostakis et al. (2015) apply the IVX methodology to examine whether stock market returns are predictable by lagged financial variables. Phillips and Lee (2016) explore inference under local explosiveness and mildly explosive roots, Demetrescu and Rodrigues (2022) introduce a bias-corrected IVX estimator, analogous to the finite-sample correction by Amihud and Hurvich (2004), Demetrescu et al. (2023) propose refinements for greater robustness, and Yang et al. (2020) develop the IVX-AR test to account for serial correlation and heteroskedasticity in the regression errors. On the other hand, Lee (2016) and Fan and Lee (2019) extend the IVX methodology to quantile predictability, addressing an important and previously underdeveloped area. The IVX approach offers valuable insights for addressing persistence in predictive regressions. Its implementation, however, involves selection of a tuning parameter used to construct instruments, and the convergence rates depends on how the instruments are formed. Lee (2016) provides a practical rule for choosing the tuning parameter in the context of quantile predictive regressions, although the precise impact of this choice on test performance remains an open question. As noted by Yang et al. (2021), the IVX framework typically assumes a zero intercept in the autoregression, which may limit flexibility in some applications.

The third strand of research applies the empirical likelihood (EL) method to predictive regression settings. Originally introduced by Owen (1988, 1990), the method offers several desirable advantages. It is nonparametric, achieves fast convergence rates, does not require the choice of tuning parameters, and delivers test statistics with a standard chi-squared limiting distribution (Wilks' theorem) regardless of the persistence level of the predictor. These properties make it attractive for predictive regressions.

Building on these advantages, the EL approach was first applied successfully in predictive mean settings by Zhu et al. (2014) and has since been extended to accommodate more complex features. For example, Li et al. (2017) allow for autocorrelated errors, Liu et al. (2019) incorporate first differences of the predictor, and Yang et al. (2021) permit lagged dependent variables. Nevertheless, despite these advances, existing EL methods have not been extended to accommodate mildly integrated or mildly explosive predictors, which are particularly relevant in many applications in economics and finance; see, for example, Phillips et al. (2011). This paper aims to fill this gap by extending the EL framework to predictive regressions that accommodate mildly integrated and mildly explosive predictors and relax regularity conditions, while also incorporating distributional characteristics beyond the mean, in particular, quantile-based inference.

In applying the EL method, an important practical issue is the presence of an unknown intercept in the predictive regression. One way to address this is through the sample-splitting (large-lag differencing) approach proposed by Zhu et al. (2014), which facilitates straightforward elimination of the intercept. We revisit and extend this approach by substantially relaxing the restrictive regularity conditions commonly imposed in the existing EL literature such as i.i.d. errors, and by accommodating additional persistence regimes for the predictor. However, in mean predictive regression settings, sample splitting sacrifices efficiency because the effective sample size is halved, and in the context of quantile predictive regressions, it is not applicable, since the differenced score no longer yields pivotal EL constraints. To complement sample splitting, we propose a new two-stage procedure for mean predictive regressions, motivated by Cai and Wang (2014), and also apply it to quantile settings (where sample splitting is infeasible). Our method retains the full sample by first estimating the intercept and then applying the EL test to the intercept-adjusted series. While effective, this procedure can suffer from small-sample distortions, especially under high persistence. We address these challenges by incorporating gradually saturated, hyperbolic tangent-type, weight functions into the estimating equations to stabilize the EL constraint and reduce higher-order bias. Simulation evidence shows that the proposed EL procedures deliver competitive size and power across persistence regimes, with notable gains in quantile predictability.

This paper makes several contributions to the existing literature. First, we develop an EL framework that accommodates a wide spectrum of predictor persistence and delivers a unified chi-square limiting theory, and apply it to both mean and quantile-based settings. Second, we develop two complementary approaches for handling unknown intercepts: a sample-splitting method under relaxed conditions, and a new two-stage procedure that improves efficiency and enables quantile inference. Third, we propose practical remedies for finite sample bias: Bartlett-type bias correction and gradually saturated weights, which enhance finite-sample performance under strong persistence. These contributions provide robust inference tools for predictive regressions under realistic persistence scenarios.

While our contributions advance the econometric literature, they also address issues of practical significance in finance and risk management. Testing predictability in a quantile setting is particularly important and attractive for several reasons. From a financial perspective, investors' decisions depend on more than just the mean and variance of returns; higher-order moments and tail behavior play a critical role in portfolio choice and risk management (see Harvey and Siddique, 2000; Dittmar, 2002; Cenesizoglu and Timmermann, 2012). From an econometric perspective, quantile regressions are well suited for skewed or heavy-tailed distributions, a common feature of financial data, and are widely used in risk management applications, such as Value-at-Risk estimation. Moreover, predictive quantile regressions avoid to a large extent the theoretical challenges associated with mean regressions, such as order-imbalance issues when

regressors are highly persistent; see, Phillips (2015).

An empirical application to U.S. equity returns demonstrates the practical utility of the proposed framework in uncovering informative dynamics of predictability. We find that mean predictability is modest and concentrated in a small set of economically relevant variables, such as inflation and short-term interest rates, consistent with the mixed evidence reported in the literature on mean predictability. By contrast, quantile-based inference reveals substantially richer dynamics across the return distribution. In particular, predictors related to interest rates and inflation exhibit stronger signals in the lower tails, while the default yield spread dominates the upper tail. These results highlight the value of moving beyond mean-based approaches to capture distributional features that are relevant for portfolio allocation, risk management, and policy evaluation.

The rest of the paper is organized as follows. Sections 2 and 3 present the proposed methods and asymptotic theory for mean and quantile predictability, respectively. Section 4 studies finite-sample behavior and discusses practical remedies, including the choice of weights and Bartlett-type bias correction. Sections 5 and 6 report simulation results and the empirical application, respectively, and Section 7 concludes. All proofs are in the Appendix; the Internet Appendix contains further simulation results. We take all stochastic processes considered in this paper equipped with the same probability measure $\mathbb{P}$. As for notations, we denote by $\Longrightarrow$, $\xrightarrow{st}$, $\xrightarrow{d}$ and $\xrightarrow{p}$ weak convergence in the Skorohod space $\mathcal{D}[0,1]$, stable convergence in law, convergence in distribution, and convergence in probability, respectively. The term stationarity means strict stationarity, rather than weak stationarity. The mathematical expression for divergence in probability; i.e., $X_t \xrightarrow{p} +\infty$ as $t \to \infty$, is understood to mean $\mathbb{P}(X_t > r) \to 1$ for every $r > 0$.

2 Empirical Likelihood for Mean Predictability

We study a widely used predictive regression model for the conditional mean; see Phillips (2015) and references therein. To fix ideas, let Y_t denote the excess return on a broad market portfolio, and let X_{t-1} denote a lagged predictor such as the dividend-price ratio, earnings yield, or an interest-rate variable. More generally, the setup can be applied to any scalar response and possibly persistent predictor in a predictive regression.

The data-generating process for $\{(Y_t, X_t)\}_{t\in\mathbb{Z}^+}$ is given by

$$Y_t = \alpha + \beta X_{t-1} + U_t, \tag{1}$$

$$X_t = \theta + \rho X_{t-1} + \varepsilon_t, \tag{2}$$

where U_t and ε_t are error terms satisfying regularity conditions that are specified shortly. The slope coefficient β captures the predictive content of X_{t-1} for Y_t. The main object of interest

is to test the null hypothesis of no mean predictability

$$H_0 : \beta = 0. \tag{3}$$

In this context, we develop an empirical likelihood (EL) approach that delivers statistical inference on β under a wide range of persistence scenarios for the predictor X_t. In what follows, we first outline the EL formulation for mean regressions and then present its asymptotic properties and implementation details.

We make the following assumptions about U_t and ε_t, which are sufficiently general to accommodate a wide range of empirical features commonly observed in practice. The proposed characterization allows for correlation between the error terms U_t and ε_t, thereby accommodating the potential embedded endogeneity in predictive regressions (Stambaugh, 1999; Campbell and Yogo, 2006).

Assumption 1. *The error terms (U_t, ε_t) are characterized as follows:*

$$U_t = \phi V_t + z_t, \quad \text{and} \quad \varepsilon_t = V_t, \tag{4}$$

where (z_t, V_t) is a martingale difference with respect to the natural filtration $\mathcal{G}_{t-1} = \sigma(\{z_s, V_s; s \leq t-1\})$ and satisfies $\sup_t \mathbb{E}(|V_t|^{2+q} + |z_t|^{2+q}) < \infty$ for some $q > 0$. Moreover, ε_t is either conditionally homoscedastic, or conditionally heteroscedastic with conditional variance ς_t^2 that is stationary and satisfies $\mathbb{E}(\varsigma_t^{-4}) < \infty$. The process (z_t, V_t) is α-mixing with rate $\alpha(\ell) \leq C\gamma^\ell$ for some $\gamma \in (0,1)$ and $C > 0$.

Remark 1. The above setup is notably weaker than what is commonly adopted in the predictive regressions literature along several dimensions. First, we substantially relax the i.i.d. condition on the error terms (U_t, V_t) often postulated in the empirical likelihood framework (see Cai et al., 2015; Li et al., 2017; Liu et al., 2019; Yang et al., 2021). Second, the intercept θ in (2) is allowed to be non-zero in the model. Lastly, we accommodate conditional heteroscedasticity in the autoregressive error ε_t, following the conditions in Lee (2018), which are satisfied by many GARCH-type processes. We note for later reference that the α-mixing assumption for (z_t, V_t) is only needed in Theorem 2. The required mixing rate can be relaxed to the polynomial rate $\alpha(\ell) \leq C\ell^{-\kappa}$ with exponent $\kappa > (2+q)/(2+2q)$ without affecting the asymptotic theory of the paper; see the proofs for details.

Assumption 2. *Further to Assumption 1, with respect to $\mathcal{G}_{t-1}$, the error term U_t is either*

(i) Conditionally homoscedastic:

$$\textit{i.e.} \qquad \mathbb{E}(U_t^2 \mid \mathcal{G}_{t-1}) = \sigma_U^2 < \infty,$$

or

(ii) Conditionally heteroscedastic:

$$i.e. \qquad \mathbb{E}(U_t^2 \mid \mathcal{G}_{t-1}) \;=\; \sigma_t^2$$

where σ_t^2 is $\mathcal{G}_{t-1}$-measurable and is finite for all t, and, as $T \to \infty$ satisfies

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(U_t^2 \mid \mathcal{G}_{t-1}) \xrightarrow{p} \sigma^2. \tag{5}$$

Remark 2. Assumption 2(ii) is mild and flexible and does not restrict the conditional variance of U_t to be stationary. Similar conditions appear in Assumption A2 of Cai et al. (2023). See also Park (2002) for related discussions. The assumption accommodates various forms of nonstationarity, such as processes with structural breaks in the volatility. In many financial applications, however, conditional volatility exhibits strong persistence and may behave as nearly integrated, as documented by Jacquier et al. (1994). Such cases typically violate Assumption 2(ii). To address this, Choi et al. (2016) propose a transformation that ensures a constant conditional variance as in (5). Practitioners seeking to handle nearly integrated volatility may find their procedure useful.

To capture a broad range of predictor persistence, we follow the standard practice of parameterizing the autoregressive coefficient in equation (2) as $\rho = \rho_T$, and let

$$\rho_T \;=\; 1 + \frac{c}{T^a},$$

where c and $a \geq 0$ are some fixed constants; see Phillips (2015) and references therein. The time-series properties of X_t are determined by the pair (c, a). In this paper, we consider the following five cases:

C1 : $|1 + c| < 1$ and $a = 0$; X_t is stationary;

C2 : $c < 0$ and $a \in (0, 1)$; X_t is mildly integrated;

C3 : $c \neq 0$ and $a = 1$; X_t is nearly integrated;

C4 : $c = 0$; X_t is integrated;

C5 : $c > 0$ and $a \in (0, 1)$; X_t is mildly explosive;

These cases encompass a wide range of potential regimes for X_t encountered in practice. Notably, within the context of EL methods, existing studies have not covered the mildly integrated

C2 and mildly explosive C5 scenarios. By contrast, the IVX literature accommodates these cases; see, for example, Phillips and Lee (2013). Together with our relaxed error assumptions (Assumptions 1 and 2), this broadens the scope of EL-based inference for predictive regressions while retaining its practical advantages, such as fast convergence rates, no tuning parameters, and ease of implementation.

2.1 When α is Known

We first consider the case where the intercept $\alpha = \alpha_0$ is known *a priori*. This assumption simplifies the EL formulation and helps illustrate the core idea underlying the method. Moreover, presenting results under this scenario provides a benchmark for the small-sample performance of EL tests, as it reflects the advantage that can be gained when the intercept is known or estimated with high precision.

Given the predictive regression (1), consider the following weighted estimating equation:

$$\sum_{t=1}^{T} (Y_t - \alpha_0 - \beta X_{t-1})\, w\,(X_{t-1}) = 0,$$

where $w(\cdot)$ is a measurable weight function. The weight function is introduced to accommodate the potential nonstationarity of X_t. Under nonstationarity, X_t may diverge in probability, which can invalidate standard EL limit theory unless appropriately weighted; see Ling (2005) and Zhu et al. (2014).

Remark 3. The specification of $w(\cdot)$ is of both theoretical and practical importance. Later in Section 4, we report some new findings that, to our knowledge, have not been previously documented in the EL-based predictive regressions literature. A set of sufficient conditions ensuring the validity of our limit theory is that the weight function is continuous, uniformly bounded, and satisfies the saturation condition $w(x)^2 \to 1$ as $|x| \to \infty$. Hereafter, we suppose that the weight function we consider satisfies these conditions.

Some examples of such weight functions include

$$w_1(x) = \frac{x}{\sqrt{1+x^2}}, \quad w_2(x) = \frac{x}{1+|x|}, \quad w_3(x) = \tanh(x/b), \tag{6}$$

for $b > 0$. In the literature, the widely adopted practice has been to employ $w_1(x)$ (see, inter alia, Zhu et al., 2014; Liu et al., 2019; Yang et al., 2021), which we will refer to as the conventional weight hereafter. In Section 4, we discuss finite-sample issues that may arise in this context, and propose alternative choices of weights, explaining how the issues can be alleviated, while preserving the validity of the asymptotic theory.

The empirical likelihood method proceeds as follows. For $t = 1, 2, \ldots, T$, we define

$$Z_t(\beta) \;=\; [Y_t - \alpha_0 - \beta X_{t-1}]\, w\left(X_{t-1}\right).$$

The profile empirical likelihood function for β is given by:

$$L_T\left(\beta\right) \;=\; \sup\left\{ \prod_{t=1}^{T} \left(T\pi_t\right) :\;\; \pi_t \geq 0 \text{ for all } t,\; \sum_{t=1}^{T} \pi_t = 1, \quad \sum_{t=1}^{T} \pi_t Z_t(\beta) = 0 \right\}, \tag{7}$$

where the supremum is taken over the π's. Accordingly, (7) defines a nonparametric analogue of the likelihood ratio; see Owen (1988, 1990). Applying the method of Lagrange multipliers to incorporate these constraints yields the empirical likelihood ratio statistic:

$$\ell_T\left(\beta\right) \;=\; -2\log L_T\left(\beta\right) \;=\; 2\sum_{t=1}^{T} \log\left\{1 + \lambda Z_t\left(\beta\right)\right\},$$

where the multiplier $\lambda = \lambda(\beta)$ satisfies

$$\sum_{t=1}^{T} \frac{Z_t\left(\beta\right)}{1 + \lambda Z_t\left(\beta\right)} = 0.$$

Theorem 1. *Suppose that the data is generated according to the process in equations (1)-(2), with the predictive variable belonging to either class C1, C2, C3, C4, or C5. Suppose also that Assumption 1 and either Assumption 2-(i) or 2-(ii) are satisfied. Then,*

$$\ell_T(\beta_0) \xrightarrow{\;d\;} \chi^2_1$$

as $T \to \infty$, where β_0 denotes the true value of β.

The theorem establishes the limit theory for the test statistic uniformly across cases C1$-$C5. As a consequence, we would reject the hypothesis $H_0 : \beta = \beta_0$ at level ϱ if $\ell_T(\beta_0) > \chi^2_{1,1-\varrho}$. Furthermore, an empirical likelihood confidence set for β can be obtained.

2.2 When α is Unknown: (i) Sample Splitting

Since the intercept α is typically unknown in practice, the conventional approach in the EL literature is to eliminate it via large-lag differencing (i.e. sample-splitting), as originally suggested by Zhu et al. (2014). This method partitions the sample and constructs differenced observations using a sufficiently large lag, thereby removing the intercept from the regression. While straightforward to implement, its main drawback is that it sacrifices efficiency because the effective sample size is halved. Nevertheless, sample-splitting offers two advantages: it mitigates the degree of nonstationarity in X_t through differencing and avoids the need to estimate the

intercept. Below we extend the sample-splitting method by (i) accommodating mildly integrated and mildly explosive predictors and (ii) substantially relaxing restrictive i.i.d. error assumptions.

In the next subsection, Section 2.3, we propose a new two-stage method as a complementary approach that exploits the full sample. Both procedures are useful in their own right and are studied within the same unified asymptotic EL framework.

Let $m = \lfloor T/2 \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function, and define

$$Y_t^* = Y_{t+m} - Y_t, \quad X_t^* = X_{t+m} - X_t, \quad \text{and} \quad U_t^* = U_{t+m} - U_t$$

for $t = 1, \ldots, m$, hence sample-splitting. By construction, the intercept α cancels out and is eliminated from the regression specification. Sample splitting preserves the property that under nonstationarity X_t^* diverges in probability when X_t is, whereas simple first differencing need not. Consequently, the empirical likelihood procedure via weighting can be carried over to the differenced regression.

The data generating process can be written in terms of the differenced variables as

$$Y_t^* = \beta X_{t-1}^* + U_t^*, \quad \text{and} \quad X_t^* = \rho X_{t-1}^* + \varepsilon_t^* \tag{8}$$

for $t = 1, \ldots, m$, and $\varepsilon_t^* = \varepsilon_{t+m} - \varepsilon_t$. In Theorem 2 below, we show that Wilks' theorem established in Theorem 1 continues to hold regardless of the degree of persistence of X_t.

The profile empirical likelihood function for β is given by:

$$L_T^*(\beta) = \sup \left\{ \prod_{t=1}^{m} (m\pi_t) : \ \pi_1 \geq 0, \ldots, \pi_m \geq 0, \ \sum_{t=1}^{m} \pi_t = 1, \ \sum_{t=1}^{m} \pi_t Z_t^*(\beta) = 0 \right\},$$

where $Z_t^*(\beta) = [Y_t^* - \beta X_{t-1}^*] w(X_{t-1}^*)$. For example, if the conventional weight $w_1(\cdot)$ in (6) is used, then $w(X_{t-1}^*) = X_{t-1}^* / \sqrt{1 + X_{t-1}^{*,2}}$. As before, using the method of Lagrange multipliers we have

$$\ell_T^*(\beta) = -2 \log L_T^*(\beta) = 2 \sum_{t=1}^{m} \log \{1 + \lambda Z_t^*(\beta)\},$$

where $\lambda = \lambda(\beta)$ satisfies

$$\sum_{t=1}^{m} \frac{Z_t^*(\beta)}{1 + \lambda Z_t^*(\beta)} = 0.$$

Theorem 2. *Suppose that the data is generated according to the process in equations (1)-(2), with the predictive variable belonging to either class C1, C2, C3, C4, or C5. Suppose also that Assumption 1 and either Assumption 2-(i) or 2-(ii) are satisfied. Then,*

$$\ell_T^*(\beta_0) \xrightarrow{d} \chi_1^2$$

as $T \to \infty$, where β_0 denotes the true value of β.

As a consequence of Theorem 2, we would reject the hypothesis $H_0 : \beta = \beta_0$ at level ϱ if $\ell_T^*(\beta_0) > \chi^2_{1,1-\varrho}$. Furthermore, an empirical likelihood confidence set for β can be obtained.

Remark 4. The sample splitting approach avoids estimating the intercept and is less sensitive to finite-sample distortions owing to its mitigated effect of predictor nonstationarity.

2.3 When α is Unknown: (ii) A Two-Stage Approach

We propose a new two-stage procedure as a complementary approach to sample splitting for handling the unknown intercept, which has not been explored in the EL-based predictive regressions literature to our knowledge. While sample splitting effectively halves the sample and may entail an efficiency loss, the proposed two-stage approach retains the full sample for inference and can yield higher efficiency.

The procedure first estimates the intercept α in the style of Cai and Wang (2014). Specifically, in the first step, we run a first order autoregression for X_t to obtain the residuals $\widehat{\varepsilon}_t$. In the second step, in view of equations (1)−(4), we regress Y_t on $\widehat{\varepsilon}_t$ and X_{t-1} to estimate α in (1) via OLS. We denote the resulting estimate by $\widetilde{\alpha}$. Consequently, we redefine the dependent variable as $\widetilde{Y}_t = Y_t - \widetilde{\alpha}$, and apply the empirical likelihood method directly to the intercept-adjusted observations $\widetilde{Y}_t$. Cai and Wang (2014) imply that $\widetilde{\alpha}$ is $\sqrt{T}$-consistent. Therefore, a non-negligible asymptotic term may arise in the limit under the $\sqrt{T}$-scaling. We propose a solution to address the issue by employing the centered weight in the two-stage approach:

$$w^c(x) \;:=\; w(x) \;-\; \frac{1}{T}\sum_{s=1}^{T} w(X_{s-1}), \tag{9}$$

in order to let the error term vanish. The centered weight function remains uniformly bounded, and the sample average of $(w^c(X_{t-1}))^2$ admits a finite probability limit. These are sufficient for the limit theory to hold; see the proof for details. We let

$$\widetilde{Z}_t(\beta) \;=\; \big[\widetilde{Y}_t - \beta X_{t-1}\big] \cdot w^c(X_{t-1}), \tag{10}$$

and the profile empirical likelihood for β is given by

$$\widetilde{L}_T\left(\beta\right) = \sup\left\{ \prod_{t=1}^{T} (T\pi_t) : \pi_1 \geq 0, \ldots, \pi_T \geq 0,\; \sum_{t=1}^{T} \pi_t = 1,\; \sum_{t=1}^{T} \pi_t \widetilde{Z}_t\left(\beta\right) = 0 \right\}.$$

The corresponding empirical likelihood ratio statistic is then

$$\widetilde{\ell}_T(\beta) \;=\; -2\log \widetilde{L}_T\left(\beta\right) \;=\; 2\sum_{t=1}^{T} \log\Big(1 + \widetilde{\lambda}\widetilde{Z}_t(\beta)\Big), \tag{11}$$

where $\widetilde{\lambda} = \widetilde{\lambda}(\beta)$ is the Lagrange multiplier satisfying:

$$\sum_{t=1}^{T} \frac{\widetilde{Z}_t(\beta)}{1 + \widetilde{\lambda}\widetilde{Z}_t(\beta)} = 0. \tag{12}$$

The intercept $\widetilde{\alpha}$ is treated as a nuisance parameter. The following theorem establishes that the two-stage EL statistic satisfies Wilks' theorem uniformly across cases C1−C5.

Theorem 3. *Suppose that the data is generated according to the process in equations (1)-(2), with the predictive variable belonging to either class C1, C2, C3, C4, or C5. Suppose also that Assumption 1 and either Assumption 2-(i) or 2-(ii) are satisfied. Then,*

$$\widetilde{\ell}_T(\beta_0) \xrightarrow{d} \chi^2_1$$

as $T \to \infty$, where β_0 denotes the true value of β.

Clearly, Theorem 3 implies that we would reject the hypothesis $H_0 : \beta = \beta_0$ at level ϱ if $\widetilde{\ell}_T(\beta_0) > \chi^2_{1,1-\varrho}$. Furthermore, an empirical likelihood confidence set for β can be obtained.

Remark 5. The two-stage method uses the full sample for inference, thereby circumventing the efficiency loss in sample splitting. Section 4 discusses finite-sample issues that may arise in the two-stage method and proposes remedies via a bias correction and alternative weight choices.

3 Empirical Likelihood for Quantile Predictability

We now extend the EL framework to predictive quantile regressions. Unlike the mean regression setting considered in Section 2, quantile regressions allow predictive relationships to vary across different parts of the conditional distribution, providing a richer characterization of return dynamics and tail risk; see, for example, Gonzalo and Pitarakis (2012). While empirical likelihood has been widely applied to mean predictive regressions, its adaptation to quantile settings poses additional challenges because the estimating equations differ and involve non-smooth, indicator-based, score functions. In what follows, we outline the EL formulation for quantile regressions, establish its asymptotic properties under the persistence scenarios introduced earlier, and discuss implementation details.

For a given quantile level $\tau \in (0,1)$, let $Q_{Y_t}(\tau \mid \mathcal{G}_{t-1})$ denote the conditional τ-quantile of Y_t given $\mathcal{G}_{t-1} = \sigma(\{z_s, V_s; s \leq t-1\})$. Following Lee (2016), Fan and Lee (2019), and Cai et al. (2023), we consider the specification:

$$Q_{Y_t}(\tau \mid \mathcal{G}_{t-1}) = \alpha_\tau + \beta_\tau X_{t-1}, \tag{13}$$

where X_{t-1} is the lagged predictor, with its data generating process defined in (2). The quantile

innovation is defined as

$$U_{t,\tau} := Y_t - Q_{Y_t}(\tau \mid \mathcal{G}_{t-1}). \tag{14}$$

A key feature of this model is that the coefficients α_τ and β_τ vary across quantile levels, allowing for a more informative assessment of predictor effects and tail risk. This flexibility is particularly important for capturing heterogeneity in predictive relationships, especially in the tails. The object of interest is to test the null hypothesis of no quantile predictability:

$$H_0 : \beta_\tau = 0$$

for some τ, which is similar to (3) for the mean model. As in the mean regression setting, we distinguish between two scenarios for the intercept α_τ: known and unknown. Section 3.1 presents the EL formulation when α_τ is known, which serves as a benchmark and illustrates the core idea of the method. When α_τ is unknown, large-lag differencing alters the conditional quantile structure, and the resulting EL constraints become non-pivotal, in the sense that the variance of the quantile score depends on an unknown conditional joint distribution. Consequently, the sample-splitting approach is not applicable in this framework; see Section 3.2 for details. Instead, we adopt the two-stage procedure introduced earlier for mean regressions in the quantile setting. This is the main reason why we propose the two-stage estimation procedure for mean regressions in Section 2.3. We continue to accommodate the persistence regimes described in cases C1−C5. We also maintain Assumption 1 to keep the exposition focused and the framework unified, although the quantile results only rely on the relevant parts.

3.1 When α_τ is Known

We first consider the case where the intercept $\alpha_\tau = \alpha_{\tau,0}$ is known *a priori*. As before, this scenario serves as a benchmark, illustrating the core idea of the EL approach and representing the potential size and power the test can achieve when the intercept is known or estimated with high precision.

Given the predictive model specified in (13), similar to the empirical likelihood approach for quantile regression as in Otsu (2008) and Wang and Zhu (2011), we consider the following weighted estimating equation:

$$\sum_{t=1}^{T} \xi_{t,\tau}(\beta_\tau) = 0,$$

where $\xi_{t,\tau}(\beta_\tau) = \psi_\tau(Y_t - \alpha_{\tau,0} - \beta_\tau X_{t-1})w(X_{t-1})$ and $\psi_\tau(u) = \tau - 1(u < 0)$ is the quantile score, which is a generalized derivative of the check function $\vartheta(u) = u(\tau - 1(u < 0))$. Then, the

profile empirical likelihood function for β_τ is

$$L_{T,\tau}(\beta_\tau) = \sup\left\{ \prod_{t=1}^{T} (T\pi_t) : \ \pi_1, \ldots, \pi_T \geq 0, \ \sum_{t=1}^{T} \pi_t = 1, \ \sum_{t=1}^{T} \pi_t \xi_{t,\tau}(\beta_\tau) = 0 \right\}.$$

Consequently, the empirical likelihood ratio statistic is given by:

$$\ell_{T,\tau}(\beta_\tau) \ = \ -2\log L_{T,\tau}(\beta_\tau) \ = \ 2\sum_{t=1}^{T} \log\{1 + \lambda_\tau \xi_{t,\tau}(\beta_\tau)\},$$

where the multiplier $\lambda_\tau = \lambda_\tau(\beta_\tau)$ satisfies

$$\sum_{t=1}^{T} \frac{\xi_{t,\tau}(\beta_\tau)}{1 + \lambda_\tau \xi_{t,\tau}(\beta_\tau)} = 0.$$

Remark 6. By definition, the quantile innovation $U_{t,\tau}$ in (14) satisfies $\mathbb{P}(U_{t,\tau} < 0 | \mathcal{G}_{t-1}) = \tau$. Hence, $\psi_\tau(U_{t,\tau}) = \tau - 1\{U_{t,\tau} < 0\}$ obeys

$$\mathbb{E}\left[\psi_\tau(U_{t,\tau}) | \mathcal{G}_{t-1}\right] = 0, \quad \text{and} \quad \mathbb{E}\left[\psi_\tau(U_{t,\tau})^2 | \mathcal{G}_{t-1}\right] = \tau(1-\tau),$$

so that $\psi_\tau(U_{t,\tau})$ forms a martingale difference sequence with respect to $\mathcal{G}_{t-1}$. Therefore, we can show that the EL constraints remain valid, and the EL procedure developed for mean regressions extends naturally to the quantile setting.

Theorem 4. *Suppose the data is generated according to the process in equations* (13) *and* (2)*, and the predictive variable belongs to either class C1, C2, C3, C4, or C5. Suppose also that Assumption 1 holds. Then, for each* $\tau \in (0,1)$ *we have*

$$\ell_{T,\tau}(\beta_{\tau,0}) \ \xrightarrow{d} \ \chi^2_1$$

as $T \to \infty$*, where* $\beta_{\tau,0}$ *denotes the true value of* β_τ*.*

From Theorem 4, one can see that for each $\tau \in (0,1)$ we would reject the hypothesis $H_0 : \beta_\tau = \beta_{\tau,0}$ at level ϱ if $\ell_{T,\tau}(\beta_{\tau,0}) > \chi^2_{1,1-\varrho}$. Furthermore, an empirical likelihood confidence set for β_τ can be obtained.

3.2 When α_τ is Unknown

We now turn to the case where the intercept α_τ is unknown. As discussed earlier, the conventional sample-splitting approach used in the EL literature for mean regressions is not directly applicable in the quantile setting. This phenomenon is similar to the case that the profile least squares method for semiparametric mean regressions is not directly applicable to semiparametric quantile models; see, for example, Cai and Xiao (2012) for details. What

happens is that large-lag differencing modifies the conditional quantile structure and leads to estimating equations whose variance depends on an unknown conditional joint distribution. To see this, as before, write $m = \lfloor T/2 \rfloor$ and define $Y_t^* = Y_{t+m} - Y_t$ and $X_t^* = X_{t+m} - X_t$ for $t = 1, \ldots, m$. Let $U_{t,\tau}^* = U_{t+m,\tau} - U_{t,\tau}$, where $U_{t,\tau} := Y_t - Q_{Y_t}(\tau \mid \mathcal{G}_{t-1})$. The conditional second moment of the score $\mathbb{E}(\psi_\tau(U_{t,\tau}^*)^2|\mathcal{G}_{t-1})$ depends on $\mathbb{P}(U_{t,\tau}^* < 0|\mathcal{G}_{t-1})$, where

$$\mathbb{P}\left(U_{t,\tau}^* < 0|\mathcal{G}_{t-1}\right) \;=\; \iint 1\{u' < u\}\, dF_{U_{t+m,\tau},U_{t,\tau}|\mathcal{G}_{t-1}}(u', u),$$

and $F_{U_{t+m,\tau},U_{t,\tau}|\mathcal{G}_{t-1}}$ is the conditional joint distribution function of $U_{t+m,\tau}$ and $U_{t,\tau}$. Thus, the variance of the differenced score depends on the unknown conditional joint distribution. Furthermore, $\psi_\tau(U_{t,\tau}^*)$ is no longer a martingale difference. The resulting EL constraint is hence not pivotal, implying that sample-splitting cannot be applied in the quantile framework.

Therefore, we instead adapt the two-stage procedure introduced earlier for mean regressions to the quantile setting. Due to the presence of quantile score, we additionally assume the following local smoothness condition around the τ-th conditional quantile. Along with weight centering, the assumption keeps the EL statistic pivotal, yielding the desired limit result. The condition is satisfied by some regime-switching models and time-varying monotone distortions. Similar positivity and Lipschitz-type conditions on the conditional density at the quantile are assumed in the quantile regression literature; see, for example, Assumption 3.1 of Otsu (2008) and Condition S.2 of Belloni et al. (2019).

Assumption 3. *There exist some constants $f_\tau(0) \in (0, \infty)$, $L_\tau < \infty$ and $\varepsilon_0 > 0$ such that for all t, the conditional density $f_{t,\tau}(0|\mathcal{G}_{t-1}) = f_\tau(0) \in (0, \infty)$, we have $|f_{t,\tau}(u \mid \mathcal{G}_{t-1}) - f_{t,\tau}(0|\mathcal{G}_{t-1})| \leq L_\tau|u|$ for all $|u| \leq \varepsilon_0$.*

Now, we consider the quantile estimator as in Koenker and Bassett (1978) and Lee (2016):

$$\left(\widetilde{\alpha}_\tau, \widetilde{\beta}_\tau\right)' := \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{t=1}^{T} \vartheta_\tau \left(Y_t - \alpha_\tau - \beta_\tau X_{t-1}\right),$$

and define

$$\widetilde{\xi}_{t,\tau}(\beta_\tau) \;=\; \psi_\tau(Y_t - \widetilde{\alpha}_\tau - \beta_\tau X_{t-1}) w^c(X_{t-1}).$$

where $w^c(\cdot)$ is the centered weight defined in (9) and $\vartheta(u) = u(\tau - 1(u < 0))$. The profile empirical likelihood function for β_τ is then

$$\widetilde{L}_{T,\tau}(\beta_\tau) = \sup \left\{ \prod_{t=1}^{T} (T\pi_t) : \; \pi_1, \ldots, \pi_T \geq 0, \; \sum_{t=1}^{T} \pi_t = 1, \; \sum_{t=1}^{T} \pi_t \widetilde{\xi}_{t,\tau}(\beta_\tau) = 0 \right\}.$$

Consequently, the empirical likelihood ratio statistic is given by:

$$\widetilde{\ell}_{T,\tau}(\beta_\tau) \;=\; -2\log \widetilde{L}_{T,\tau}(\beta_\tau) \;=\; 2\sum_{t=1}^{T}\log\{1+\lambda_\tau \widetilde{\xi}_{t,\tau}(\beta_\tau)\},$$

where the multiplier $\lambda_\tau = \lambda_\tau(\beta_\tau)$ satisfies

$$\sum_{t=1}^{T}\frac{\widetilde{\xi}_{t,\tau}(\beta_\tau)}{1+\lambda_\tau \widetilde{\xi}_{t,\tau}(\beta_\tau)} = 0. \tag{15}$$

Theorem 5. *Suppose the data is generated according to the process in equations (13) and (2), and the predictive variable belongs to either class C1, C2, C3, C4, or C5. Suppose also that Assumption 1 and 3 hold. Then, for each $\tau \in (0,1)$ we have*

$$\widetilde{\ell}_{T,\tau}(\beta_{\tau,0}) \xrightarrow{d} \chi^2_1$$

as $T \to \infty$, where $\beta_{\tau,0}$ denotes the true value of β_τ.

As a consequence of Theorem 5, for each $\tau \in (0,1)$ we would reject the hypothesis $H_0 : \beta_\tau = \beta_{\tau,0}$ at level ϱ if $\widetilde{\ell}_{T,\tau}(\beta_{\tau,0}) > \chi^2_{1,1-\varrho}$. Furthermore, an empirical likelihood confidence set for β_τ can be obtained.

4 Bias Correction and Weight Choices

This section documents a potential finite-sample bias in the proposed two-stage EL method. We explain why the bias can be particularly pronounced when the regressor is highly persistent. As we show below, this is so even though EL ratio statistic admits the Wilks-type chi-squared limit as in the previous sections. The Taylor expansion of (12) (equivalently, in the quantile case, (15)) yields

$$0 = \frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_t - \widetilde{\lambda}\frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_t^2 + \underbrace{\widetilde{\lambda}^2\frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_t^3 - \widetilde{\lambda}^3\frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_t^4}_{=:\zeta} + \cdots.$$

Since $\widetilde{\lambda} = O_p(T^{-1/2})$, it follows that

$$\widetilde{\lambda} \;=\; \frac{\frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_t}{\frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_t^2} + \frac{\zeta}{\frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_t^2} \;=\; \frac{\frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_t}{\frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_t^2} + O_p\!\left(\frac{1}{T}\right).$$

Table 1. A simulation result for two-stage EL and the conventional weight $X_{t-1}/\sqrt{1+X_{t-1}^2}$

c	saturation	skewness	kurtosis	q/T
-50	0.088	-0.0125	3.506	0.007
-20	0.240	-0.169	4.278	0.009
0	0.968	7.900	93.431	0.104
2	0.968	7.880	91.935	0.101

By Taylor expansion to $\log(1+\widetilde{\lambda}\widetilde{Z}_t)$ in (11), it follows that

$$\begin{aligned}\widetilde{\ell}_T(\beta_0) &= 2\widetilde{\lambda}\sum_{t=1}^{T}\widetilde{Z}_t - \widetilde{\lambda}^2\sum_{t=1}^{T}\widetilde{Z}_t^2 + \frac{2\widetilde{\lambda}^3}{3}\sum_{t=1}^{T}\widetilde{Z}_t^3 - \frac{\widetilde{\lambda}^4}{2}\sum_{t=1}^{T}\widetilde{Z}_t^4 + \cdots \\ &= \frac{\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{Z}_t\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_t^2} + f_T(\zeta) + o_p(1), \qquad (16)\end{aligned}$$

where $f_T(\zeta)$ represents the influence of higher-order terms involving the third and fourth moments. Equation (16) delivers the chi-squared limit under standard regularity conditions, hence our asymptotic theorems in the previous sections are valid. However, in finite samples the higher-order terms may be non-negligible, leading to size distortions when the distribution of $\widetilde{Z}_t$ exhibits pronounced skewness and kurtosis.

4.1 Bartlett Correction

Writing $\mathbb{E}(f_T(\zeta)) = q/T$, under restrictive conditions including the i.i.d. of $\widetilde{Z}_t$, DiCiccio et al. (1991) and Liu and Chen (2010) show that the Bartlett constant q satisfies

$$q = \frac{\text{kurtosis}(\widetilde{Z}_t)}{2} - \frac{\text{skewness}(\widetilde{Z}_t)^2}{3}. \qquad (17)$$

The finite-sample bias can become pronounced when X_t is highly persistent, because skewness and kurtosis of $\widetilde{Z}_t := U_t w^c(X_{t-1}) = U_t \cdot (w(X_{t-1}) - \overline{w}(X_{t-1}))$ may be large as we explain below. As an example, Table 1 reports the sample skewness and kurtosis of $\widetilde{Z}_t$ when the conventional weight $w_1(X_{t-1}) = X_{t-1}/(1+X_{t-1}^2)^{1/2}$ is used as usual. For illustration we also report a saturation measure, defined as mean$(|w(X_{t-1})| > 0.95)$. The data are generated from the specification in equation (2), with $\theta = 0$ and $\rho_T = 1 + c/T$.

Table 1 shows that in highly persistent regimes such as integrated ($c = 0$) or mildly explosive ($c = 2$) predictors, skewness and kurtosis can be very large, leading to non-negligible values of q/T. Consequently, the test statistic (16) is inflated and finite-sample size control becomes unreliable. The mechanism is intuitive: if X_t is highly persistent, whenever it drifts from

zero it can remain large for long stretches. If the weight $w(\cdot)$ saturates quickly toward ± 1, then $|w(X_{t-1})|$ remains close to 1 for extended periods, producing a distribution with long runs near saturation and occasional extreme deviations when large shocks occur. If X_t drifts predominantly in one direction, the distribution becomes asymmetric (skewed), and long runs near saturation combined with rare jumps inflate the fourth moment (kurtosis).

Remark 7. While this phenomenon is severe for the two-stage method in 2.3, it is not pronounced under sample splitting since differencing mitigates the degree of persistence. Note that when $\theta \neq 0$ and is large, the phenomenon can be exacerbated because X_t is shifted away from zero even initially, although a small deviation from 0 does not affect the overall performance significantly, as expected from the saturation mechanism. Having a large value of non-zero ϕ also worsens the problem since $U_t = \phi V_t + z_t$; an additional component driven by V_t appears, introduces additional dependence between the EL score and the persistent regressor, amplifying size distortion. See Remark 9 in Section 5 for further discussion.

Since the Bartlett constant q involves population quantities, it needs to be estimated for a bias correction. However, under nonstationary and dependent scores, the estimation of q is not straightforward, and the true Bartlett constant may depend on the persistence of X_t and quantities beyond marginal skewness and kurtosis, so a fully reliable estimator is nontrivial. As a heuristic, we propose to consider a "naive" Bartlett bias correction:

$$\widehat{q} := \frac{\widehat{\text{kurtosis}}(\widetilde{Z}_t)}{2} - \frac{\widehat{\text{skewness}}(\widetilde{Z}_t)^2}{3} \quad \text{and} \quad \widetilde{\ell}_T(\beta_0)^{\text{naive}} := \frac{\widetilde{\ell}_T(\beta_0)}{1 + \widehat{q}/T}. \tag{18}$$

Here, we need to emphasize that the correction is a heuristic; our simulations (See Section 5) indicate that (18) can lead to a modest improvement in finite-sample size. Developing a more reliable estimator of the Bartlett constant in highly persistent regimes would be valuable, but we leave it for future work. See Kitamura et al. (2004) and references therein for more discussions.

4.2 Weight Choice

Beyond the heuristic Bartlett bias correction we discussed above, we propose modifying the weight function as a more direct and practically effective way to mitigate the finite-sample distortions. The choice of $w(\cdot)$ directly shapes the higher-order moments of the EL score and can substantially improve size control in highly persistent regimes. In particular, we find that moving beyond the conventional weight functions commonly used in the EL literature yields marked finite-sample improvements, which is confirmed by the simulation results in Section 5 (Table 3). The core idea is to let $w(\cdot)$ approach ± 1 less rapidly than the conventional weight, and hence prevents the high moments of the EL score from being inflated. Figure 1 compares three types of weight functions introduced in (6). The hyperbolic tangent weights $w_3(\cdot)$ approach

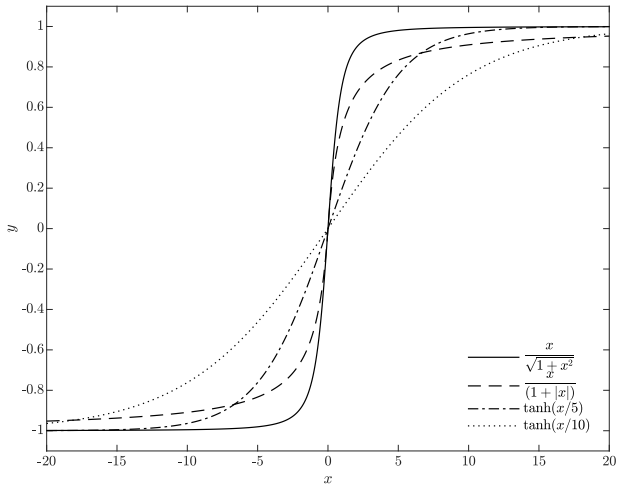


Figure 1. Plots of three weight functions

± 1 more gradually than $w_1(\cdot)$, which has served as the *de facto* conventional choice in the literature; see, inter alia, Zhu et al. (2014), Yang et al. (2021). Consequently, they mitigate clustering near saturation, reduce the inflated skewness and kurtosis, and may substantially improve finite-sample behavior, while preserving the validity of the asymptotic theory.

Table 2 extends the results in Table 1, and reports the summary statistics of different weight functions. They confirm the visual impression from Figure 1. In particular, conventional weights exhibit high saturation and extreme kurtosis when the predictor is persistent, while hyperbolic tangent weights substantially reduce such effects. For example, when $c = 0$ (unit root), saturation drops from 96.8% for the conventional weight ($w_1(\cdot)$) to 4% for the hyperbolic tangent weight with $b = 10$ (i.e. $w_3(\cdot)$). These findings motivate the use of hyperbolic tangent weights as a practical alternative to the conventional choice. They mitigate the inflation of higher-order moments that drive finite-sample bias in the two-stage EL when X is highly persistent.

Remark 8. The choice of the scale b in $w(x) = \tanh(x/b)$ is an interesting problem in practice. Setting a larger b delays saturation and reduces the clustering of $w(X_{t-1})$ near ± 1 in highly persistent regimes, thereby stabilizing the higher-order moments of the EL score. In finite samples, a less rapidly saturating weight also preserves more variation in the centered weight $w^c(X_{t-1})$. However, taking b excessively large provides little additional benefit. In our sensitivity analysis over $b = 1, \ldots, 20$ (available upon request), rejection probabilities reduce toward the nominal level as b increases, but approach a plateau around $b \approx 8$–11 when X is highly persistent, e.g. $c = 0$, after which the empirical sizes drift upward for larger b. Also, the improvement in power becomes negligible once b reaches moderately large (e.g., $b \geq 5$). Intuitively, an excessively large b makes $\tanh(x/b)$ nearly linear over a wide range of values, weakening the intended saturation-based stabilization in the EL procedure. In the spirit of choosing tuning parameters for testing

Table 2. Summary statistics of different weight functions

c	$X_{t-1}/\sqrt{1+X_{t-1}^2}$			$X_{t-1}/(1+\lvert X_{t-1}\rvert)$			$\tanh(X_{t-1}/10)$		
	Sat.	Skew.	Kurt.	Sat.	Skew.	Kurt.	Sat.	Skew.	Kurt.
-50	0.088	-0.125	3.506	0.000	-0.136	3.428	0.000	-0.826	7.293
-20	0.240	-0.169	4.278	0.000	-0.137	4.408	0.000	-0.259	7.764
0	0.968	7.900	93.431	0.384	2.504	30.696	0.042	0.948	21.448
2	0.968	7.880	91.935	0.724	2.554	30.607	0.073	1.371	29.493

Note: The table reports Monte Carlo averages of skewness (Skew.) and kurtosis (Kurt.) of the centered weight $\widetilde{Z}_t$ and those of saturation (Sat.; defined as $\mathbb{P}(|w(X_{t-1})| > 0.95)$ under different persistence levels (c). Summary statistics are reported for three weight specifications. Data for X_t are generated from the AR(1) specification in equation (2), with $\rho_T = 1 + c/T$, $\theta = 0$, and innovations drawn from a standard normal distribution. Results are based on 10,000 replications with sample size $T = 250$.

by balancing size control and power (see, for example Gao and Gijbels (2008)), we adopt $b = 10$, the middle value of the observed plateau in the sensitivity analysis, as a default. We remark that finite-sample results are very similar for nearby choices of b.

5 Simulation Study

This section examines the small-sample properties of the EL tests proposed in Sections 2 and 3. We specify the simulation setup and present evidence on the performance of the proposed methods under different persistence regimes, comparing them to existing approaches. To preserve space and maintain focus, we report here results only for both predictive mean and quantile regressions under homoscedastic errors. The simulation setup and results for heteroscedastic errors, which lead to qualitatively similar conclusions, are fully provided in the supplementary material (Internet Appendix).

5.1 Simulation Setup

In this subsection, we use Monte Carlo simulations to investigate the finite sample behavior of the proposed EL methods. For the mean regression experiments, we generate the data from the process defined in equations (1) and (2), with $\alpha = 0.2$, $\phi = 0.05$, and $c \in \{-50, -20, 0, 2\}$, corresponding to stationary, local-to-unity, unit root, and mildly explosive predictors. The autoregressive process in equation (2) is initialized at $X_0 = 0$, following the usual convention. The vector $(z_t', V_t')'$ is drawn from a bivariate standard normal distribution, and we set $\varepsilon_t = V_t$ and $U_t = \phi V_t + z_t$ as defined in (4), with $\phi \in \{-0.5, -0.20, 0\}$.[1] For quantile regressions, we use the same simulation setup for X_t, ε_t, and U_t, but generate Y_t from

$$Y_t = \alpha_\tau + \beta_\tau X_{t-1} + U_{t,\tau},$$

[1] We also considered the cases of $\phi \in \{0.5, 0.2\}$, but the results were qualitatively similar to those for $\phi \in \{-0.5, -0.2\}$. Hence, we omit them for brevity here; those results are available upon request.

Table 3. Finite-sample sizes for predictive mean regressions

ϕ	c	Conventional weights without Bartlett correction			Conventional weights with Bartlett correction			Hyperbolic tangent weights with Bartlett correction			IVX
		EL1	EL2	EL3	EL1	EL2	EL3	EL1	EL2	EL3	
0	-50	0.0533	0.0531	0.0518	0.0520	0.0508	0.0513	0.0538	0.0532	0.0543	0.0536
	-20	0.0513	0.0526	0.0519	0.0508	0.0515	0.0506	0.0524	0.0498	0.0523	0.0539
	0	0.0494	0.0572	0.0804	0.0490	0.0558	0.0739	0.0483	0.0545	0.0463	0.0469
	2	0.0491	0.0546	0.0994	0.0486	0.0526	0.0911	0.0486	0.0518	0.0512	0.0471
-0.2	-50	0.0525	0.0522	0.0526	0.0515	0.0506	0.0520	0.0526	0.0525	0.0545	0.0540
	-20	0.0502	0.0521	0.0540	0.0496	0.0505	0.0527	0.0526	0.0512	0.0541	0.0530
	0	0.0484	0.0556	0.0851	0.0480	0.0544	0.0785	0.0498	0.0537	0.0519	0.0460
	2	0.0492	0.0521	0.1040	0.0485	0.0510	0.0945	0.0498	0.0511	0.0549	0.0473
-0.5	-50	0.0501	0.0516	0.0537	0.0490	0.0501	0.0516	0.0524	0.0518	0.0561	0.0516
	-20	0.0502	0.0507	0.0576	0.0495	0.0494	0.0557	0.0543	0.0505	0.0579	0.0531
	0	0.0492	0.0544	0.1068	0.0484	0.0530	0.1007	0.0546	0.0567	0.0766	0.0439
	2	0.0495	0.0524	0.1353	0.0488	0.0513	0.1245	0.0551	0.0558	0.0780	0.0450

Note. The table reports the empirical size (i.e., the probability of incorrectly rejecting the null hypothesis of no predictability). Results are shown for: EL1 (the EL method, where α is treated as known), EL2 (sample-splitting EL approach in Section 2.2), EL3 (the two-stage EL procedure using the projection method in Section 2.3), and IVX (benchmark test of Kostakis et al. (2015) and Phillips and Lee (2016)). All tests are conducted at the 5% nominal significance level. EL tests are conducted using either the conventional weight function $X_{t-1}/(1+X_{t-1}^2)^{1/2}$, with and without the Bartlett correction in Section 4.1, or the tanh function $\tanh(X_{t-1}/10)$ with Bartlett correction. The simulation design accounts for various levels of persistence in the predictor X_t through the localizing constant $c \in \{-50, -20, 0, 2\}$, and endogeneity through the innovation correlation parameter $\phi \in \{-0.5, -0.2, 0\}$. $\theta = 0.05$. Rejection probabilities are based on 10,000 Monte Carlo simulations and sample size $T = 250$. For detailed description of the simulation design, please see Section 5.

with $U_{t,\tau} = U_t - Q_{U_t}(\tau \mid \mathcal{G}_{t-1})$, where $Q_{U_t}(\tau \mid \mathcal{G}_{t-1})$ denotes the conditional τ-quantile of U_t. All tests considered are conducted under the null hypothesis of no predictability at the 5% nominal level. We report results for $T = 250$ based on 10,000 Monte Carlo replications. This design allows us to assess the impact of persistence and endogeneity (captured by c and ϕ, respectively) on the size and power of the proposed EL tests.

5.2 Small-Sample Properties for Mean Predictability Tests

First, we evaluate the small-sample properties of the EL-based tests for mean predictability. Table 3 and Figure 2 summarize the rejection probabilities and power plots of four tests: (i) the EL procedure with the intercept α treated as known (EL1), (ii) the sample-splitting empirical likelihood approach studied in Section 2.2 (EL2), (iii) the two-stage EL procedure (EL3) in Section 2.3, where α is first estimated and removed using the consistent projection method in the style of Cai and Wang (2014), and (iv) the IVX test of Kostakis et al. (2015) and Phillips and Lee (2016), reported for benchmarking.

Several interesting conclusions emerge from these simulation results shown in Table 3. First, the finite-sample remedies in Section 4, especially the hyperbolic tangent weighting, are highly effective in improving size control in the highly persistent regimes relative to the conventional

weighting used in the EL literature. When the endogeneity level ϕ is mild, at the 5% nominal level, size is overall well calibrated for all procedures after finite-sample remedies. For EL3, this reflects the EL constraint being well behaved and the intercept being estimated precisely enough in the two-stage approach, so that the resulting score has near-zero mean under the null. In the stationary ($c = -50$) and local-to-unity ($c = -20$) cases, the power of EL3 is also close to the known-intercept upper bound (EL1), indicating that the full-sample two-stage procedure preserves most of the efficiency of the oracle test. The sample-splitting approach (EL2) also performs well and is correctly sized across all persistent levels. However, as expected, it is less powerful because splitting sacrifices information and effective sample size. Relative to IVX, EL3 is typically competitive. In the stationary and local-to-unity cases, the two-stage power curve rises at nearly the same rate as EL1 and often sits above IVX, especially near the null, where local alternatives are weak.

In the unit-root case ($c = 0$), the two-stage procedure remains competitive in power relative to IVX when endogeneity is low, and its curve converges to one slightly faster than IVX for values of β close to the null. Nevertheless, the gap between EL3 and EL1 becomes more visible, reflecting the impact of estimating α. For mildly explosive predictors ($c = 2$), the IVX curve tends to converge slightly faster than the two-stage EL, although the difference is somewhat marginal.

Remark 9. Under strong endogeneity ($\phi = -0.5$), EL3 is mildly oversized when the predictor is highly persistent, even with the gradually saturating hyperbolic tangent weights. This behavior is expected and is consistent with the construction of EL3: our two-stage adjustment removes the intercept but does not purge the innovation-driven component in the regression error, as noted in Remark 7. When $\phi \neq 0$, the EL score in (10) inherits the influence of the term $\phi V_t (= \phi\varepsilon_t)$, which is correlated with the centered weight. This correlation can lead to stronger dependence and heavier tails in the moment condition, inflating the finite-sample rejection probability.

As a remedy for this mild distortion, we consider two approaches and conduct additional diagnostic simulations. First, an "oracle purging" that removes $\phi V_t (= \phi\varepsilon_t)$ using the true innovation employs the residual $Y_t - \widetilde{\alpha} - \widetilde{\phi}\varepsilon_t$ in the EL score. This eliminates the size distortion across persistence classes, confirming that the remaining over-rejection is driven by the innovation-correlated component. Second, a "feasible purging" based on the estimated innovation can be proposed, where $Y_t - \widetilde{\alpha} - \widetilde{\phi}\widehat{\varepsilon}_t$ is used as the residual. This ameliorates the over-sizing by construction, but may introduce size distortions in stationary or near-stationary regimes due to estimation error in the autoregressive coefficient ρ (which vanishes as the sample size grows, but slowly than the highly persistent cases). These simulation results are reported in the Internet Appendix (Section IA2). Developing a feasible purging scheme that further reduces this estimation-error effect in finite samples is left for future research.

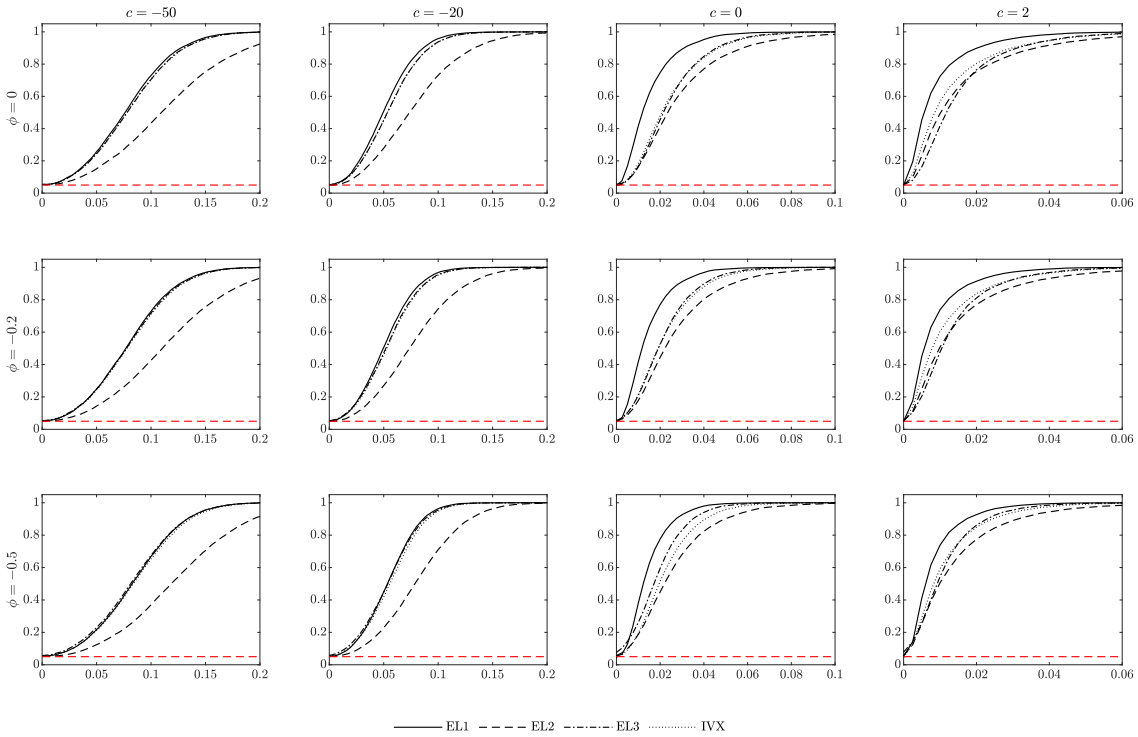


Figure 2. Finite-sample size and power plots for mean predictability tests with homoscedastic errors and tanh-based weight

Note. The figure summarizes rejection probabilities (y-axis) for tests of the null hypothesis of no predictive ability in the mean regression model (1). Results are shown for: EL1 (the EL method, where α is treated as known; Section 2.1), EL2 (sample-splitting EL approach in Section 2.2), EL3 (the two-stage EL procedure using the projection method in Section 2.3, and IVX (benchmark test of Kostakis et al. (2015) and Phillips and Lee (2016)). All EL tests are conducted using the hyperbolic tangent weight function $w(X_{t-1}) = \tanh(X_{t-1}/10)$ and applying the Bartlett correction. The x-axis represents true values of the slope coefficient β, with $\beta = 0$ corresponding to empirical size. The red dashed line marks the 5% nominal level. Rejection probabilities are based on 10,000 Monte Carlo simulations. Results are reported for sample size $T = 250$, $\theta = 0.05$, innovation correlation parameter $\phi \in \{-0.5, -0.2, 0\}$, and localizing constant $c \in \{-50, -20, 0, 2\}$. Results under different settings including heteroscedastic errors are similar, and can be found in the Internet Appendix.

Overall, the empirical likelihood procedures exhibit competitive small-sample performance for mean predictability across the persistence regimes considered. In stationary and local-to-unity cases, both procedures deliver accurate size control, while EL3 achieves higher power by exploiting the full sample. As persistence increases toward unit-root and mildly explosive settings, EL2 continues to provide steady calibration, whereas EL3 often remains competitive with IVX in terms of power, though it may exhibit mild over-rejection under strong endogeneity as discussed above. Taken together, the evidence suggests that when endogeneity is mild or controlled, the two-stage method (EL3) delivers near-oracle power with good size control, whereas under strong endogeneity and high persistence, sample splitting (EL2) can serve as a useful method that provides a stable calibration.

Table 4. Finite-sample sizes for predictive quantile regressions

c	τ	$\phi = 0$			$\phi = -0.2$			$\phi = -0.5$		
		EL1	EL3	IVX	EL1	EL3	IVX	EL1	EL3	IVX
-50	0.1	0.0578	0.0670	0.0665	0.0571	0.0692	0.0680	0.0543	0.0654	0.0634
	0.2	0.0530	0.0609	0.0521	0.0529	0.0599	0.0519	0.0522	0.0608	0.0422
	0.3	0.0524	0.0589	0.0453	0.0519	0.0590	0.0448	0.0482	0.0555	0.0445
	0.4	0.0520	0.0543	0.0410	0.0498	0.0546	0.0411	0.0489	0.0551	0.0390
	0.5	0.0486	0.0524	0.0391	0.0480	0.0520	0.0372	0.0499	0.0560	0.0362
	0.6	0.0500	0.0538	0.0397	0.0486	0.0526	0.0394	0.0466	0.0529	0.0363
	0.7	0.0514	0.0537	0.0458	0.0509	0.0530	0.0476	0.0474	0.0563	0.0398
	0.8	0.0506	0.0555	0.0535	0.0481	0.0568	0.0491	0.0505	0.0552	0.0480
	0.9	0.0595	0.0692	0.0702	0.0606	0.0652	0.0640	0.0576	0.0645	0.0603
-20	0.1	0.0541	0.0673	0.0671	0.0540	0.0653	0.0691	0.0538	0.0624	0.0640
	0.2	0.0512	0.0569	0.0553	0.0544	0.0603	0.0538	0.0520	0.0631	0.0448
	0.3	0.0513	0.0571	0.0480	0.0504	0.0564	0.0471	0.0511	0.0580	0.0422
	0.4	0.0509	0.0531	0.0461	0.0472	0.0533	0.0451	0.0513	0.0547	0.0379
	0.5	0.0476	0.0493	0.0440	0.0478	0.0515	0.0400	0.0487	0.0550	0.0348
	0.6	0.0511	0.0542	0.0390	0.0498	0.0543	0.0405	0.0461	0.0549	0.0369
	0.7	0.0505	0.0528	0.0453	0.0480	0.0555	0.0450	0.0509	0.0579	0.0406
	0.8	0.0490	0.0553	0.0528	0.0516	0.0553	0.0506	0.0500	0.0574	0.0430
	0.9	0.0572	0.0670	0.0675	0.0547	0.0638	0.0676	0.0505	0.0631	0.0630
0	0.1	0.0532	0.0525	0.1803	0.0520	0.0533	0.1717	0.0526	0.0593	0.1158
	0.2	0.0483	0.0502	0.1588	0.0505	0.0467	0.1492	0.0529	0.0558	0.0849
	0.3	0.0460	0.0452	0.1487	0.0453	0.0470	0.1361	0.0503	0.0553	0.0787
	0.4	0.0460	0.0439	0.1466	0.0457	0.0429	0.1297	0.0512	0.0568	0.0777
	0.5	0.0496	0.0423	0.1430	0.0471	0.0449	0.1287	0.0488	0.0584	0.0720
	0.6	0.0459	0.0436	0.1416	0.0487	0.0477	0.1283	0.0484	0.0611	0.0715
	0.7	0.0456	0.0475	0.1426	0.0466	0.0465	0.1357	0.0499	0.0608	0.0682
	0.8	0.0484	0.0456	0.1567	0.0493	0.0478	0.1483	0.0504	0.0634	0.0801
	0.9	0.0492	0.0582	0.1745	0.0495	0.0563	0.1754	0.0536	0.0659	0.1061
2	0.1	0.0513	0.0647	0.2130	0.0496	0.0645	0.2087	0.0530	0.0660	0.1726
	0.2	0.0488	0.0520	0.1921	0.0511	0.0519	0.1877	0.0545	0.0618	0.1375
	0.3	0.0477	0.0486	0.1791	0.0475	0.0483	0.1676	0.0527	0.0586	0.1267
	0.4	0.0487	0.0492	0.1684	0.0482	0.0468	0.1677	0.0488	0.0608	0.1176
	0.5	0.0495	0.0441	0.1652	0.0464	0.0467	0.1577	0.0507	0.0652	0.1158
	0.6	0.0470	0.0464	0.1698	0.0476	0.0513	0.1581	0.0490	0.0622	0.1193
	0.7	0.0456	0.0434	0.1730	0.0465	0.0451	0.1679	0.0505	0.0628	0.1214
	0.8	0.0476	0.0507	0.1853	0.0481	0.0508	0.1831	0.0513	0.0612	0.1319
	0.9	0.0488	0.0598	0.2167	0.0485	0.0642	0.2124	0.0512	0.0782	0.1707

Note: The table reports the empirical size (i.e., the probability of incorrectly rejecting the null hypothesis of no predictability) in the quantile regression model (13), for a selected set of quantiles $\tau \in \{0.1, 0.2, \ldots, 0.9\}$. Results are shown for EL1 (the EL method, where the intercept α_τ is treated as known; Section 3.1), EL3 (the two-stage EL procedure in Section 3.2), and IVX (benchmark test of Lee (2016)). All tests are conducted at the 5% nominal significance level. EL tests are conducted using tanh-based weights and applying the Bartlett correction. The simulation design accounts for various levels of persistence in the predictor X_t through the localizing constant $c \in \{-50, -20, 0, 2\}$, $\theta = 0.05$, and endogeneity through the innovation correlation parameter $\phi \in \{-0.50, -0.20, 0\}$. Rejection probabilities are based on 10,000 Monte Carlo simulations and sample size $T = 250$.

5.3 Small-Sample Properties for Quantile Predictability Tests

We now examine the finite-sample performance of the proposed EL-based tests for quantile predictability. As discussed in Section 3.2, the sample-splitting EL2 is not applicable to quantile inference, so we focus on EL3 here. Table 4 depicts the results for test size and Figure 3 summarizes rejection probabilities across persistence regimes and endogeneity levels, comparing

EL1 (the EL method, where the intercept α_τ is treated as known), EL3 (the two-stage EL procedure outlined in Section 3.2), and IVX (the test of Lee (2016)). To preserve clarity and readability, for Figure 3, we report results for a selected set of quantiles $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ that capture both tails and the center of the distribution as well as $\phi = -0.5$. The intercept is set to be $\theta = 0.05$ and the error is homoscedastic. Results for all other quantiles, for different levels of endogeneity $\phi = 0, -0.2$, with $\theta = 0$, and under heteroscedasticity are qualitatively similar, and our test continues to exhibit good finite sample performance. Those further simulation results can be found in the Internet Appendix.

The results of Table 4 show that the EL-based quantile tests perform very well across all persistence regimes considered, with rejection probabilities generally close to the nominal 5% level across quantiles. First, in the stationary ($c = -50$) and local-to-unity ($c = -20$) cases, at the 5% nominal level, size is well controlled for EL1 and EL3. This reflects the EL constraint being well behaved when regressors are not highly persistent and the intercept estimation error remaining negligible. In these regimes, seeing from Figure 3, the power of EL3 is virtually indistinguishable from EL1 across all reported quantiles, indicating that the two-stage procedure preserves near-oracle efficiency. IVX is competitive in these cases but generally exhibits slightly lower power than EL3, particularly near the null where local alternatives matter most.

As persistence increases, differences become more pronounced. In the unit-root case ($c = 0$), EL3 remains competitive relative to IVX for moderate signals and often converges to one faster near the null. IVX tends to over-reject in this design, especially at extreme quantiles, although its size improves somewhat as endogeneity increases while still remaining well above nominal. EL3 exhibits good size properties that are consistent across endogeneity levels, though the gap between EL3 and EL1 widens. This reflects the impact of estimating α_τ. Under high persistence, the intercept error interacts with serial dependence to tilt the EL constraint, reducing local efficiency and, when endogeneity is present, inflating size slightly at the tails. For mildly explosive predictors ($c = 2$), these patterns intensify. IVX tends to converge marginally faster than EL3 for large signals, particularly at extreme quantiles, but its size distortions are more pronounced whereas EL3 remains stable.

Overall, the two-stage EL procedure exhibits strong finite-sample properties for quantile predictability. In stationary and moderately persistent settings, it tracks the oracle EL1 closely and performs very well in terms of power. Under unit-root and mildly explosive regimes, EL3 remains competitive relative to IVX. The evidence suggests that the two-stage EL delivers good power and size across quantiles and persistence levels.

Remark 10. We note that the mild over-rejection observed for mean predictability under strong endogeneity (cf. Remark 9) is less pronounced in the quantile predictability results. This is because the quantile EL moment is built on the bounded score $\psi_\tau(u) = \tau - 1\{u < 0\}$ applied to the quantile innovation $U_{t,\tau}$. Therefore, the innovation-driven component linked to ϕV_t affects

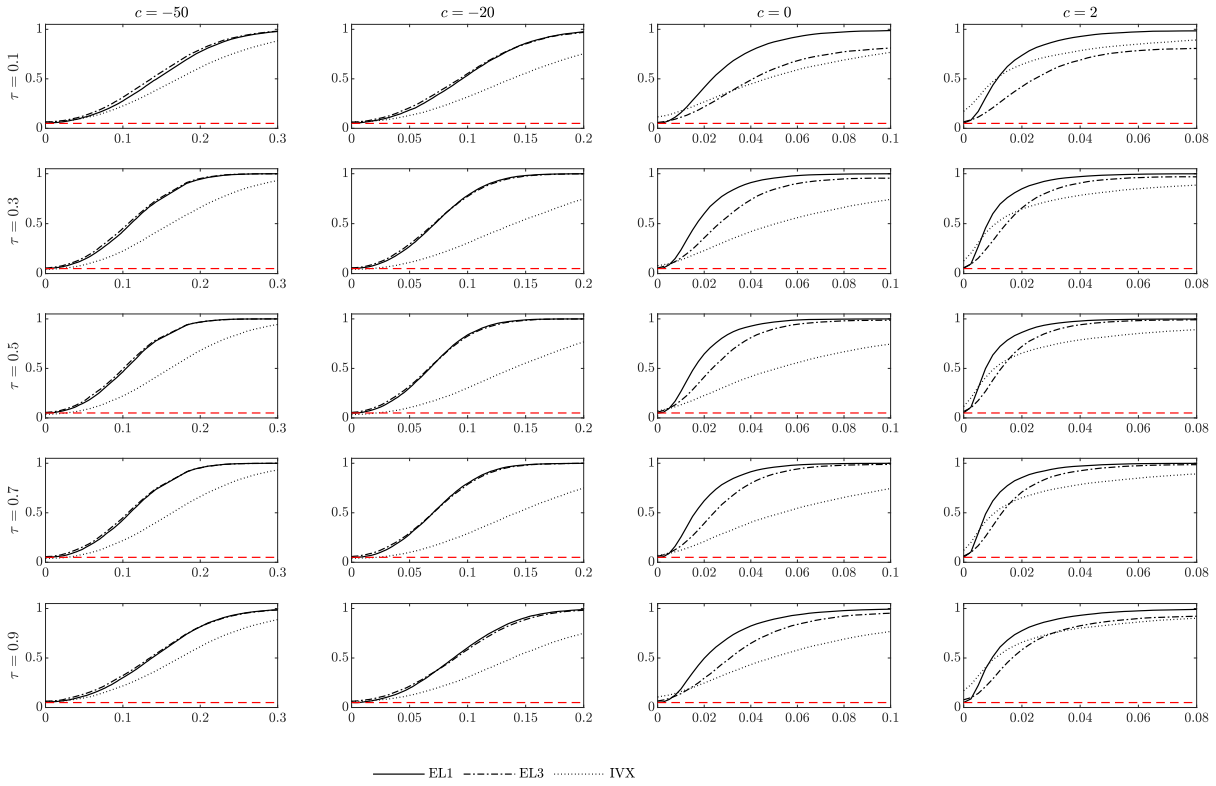


Figure 3. Finite-sample size and power plots for quantile predictability tests with homoscedastic errors and correlation $\phi = -0.5$

Note: The figure summarizes rejection probabilities (y-axis) for tests of the null hypothesis of no predictive ability in the quantile regression model (13), for a selected set of quantiles $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. Results are shown for EL1 (the EL method, where the intercept α_τ is treated as known), EL3 (the two-stage EL procedure outlined in Section 3.2), and IVX (benchmark test of Lee (2016)). EL tests are conducted using tanh-based weights with Bartlett correction. The x-axis represents true values of the slope coefficient β_τ, with $\beta_\tau = 0$ corresponding to empirical size. The red dashed line marks the 5% nominal level. Rejection probabilities are based on 10,000 Monte Carlo simulations (see Section 5 for details of the simulation design). Results are reported for sample size $T = 250$, localizing constant $c \in \{-50, -20, 0, 2\}$, $\theta = 0.05$, and innovation correlation parameter $\phi = -0.5$.

the moment primarily through the sign of the quantile residual via the indicator function rather than its magnitude. Consequently, increases in $|\phi|$ are less likely to generate the heavy-tail and higher-moment inflation of the EL score that can arise in the mean case, where the moment depends linearly on $U_t = \phi V_t + z_t$. This helps stabilize the EL constraint under high persistence and yields more reliable finite-sample calibration across quantiles.

To summarize the findings of this section, the simulation evidence for both mean and quantile predictability indicates that the proposed EL procedures offer strong finite-sample performance, with well-controlled size overall, while remaining competitive with IVX under high persistence. We now turn to an empirical application to assess how these properties translate into real-world predictability patterns.

Table 5. Empirical results for predictive mean regressions

Variable	EL2	EL3	IVX
Dividend price ratio	0.93	0.21	0.29
Dividend yield	0.88	0.16	0.24
Earnings price ratio	0.79	0.46	0.32
Dividend payout ratio	0.19	0.29	0.07
Book-to-market ratio	0.85	0.66	0.53
Net equity expansion	0.53	0.32	0.34
Treasury bill rate	0.04	0.03	0.02
Long term yield	0.11	0.11	0.08
The term spread	0.24	0.12	0.15
Default yield spread	0.66	0.29	0.30
Inflation	0.01	0.03	0.01

Note: This table reports p-values for the null hypothesis of no mean return predictability. The dependent variable is the monthly continuously compounded return on the CRSP value-weighted index in excess of the one-month Treasury bill rate. Predictive variables are listed in the first column. Results are based on EL2 (the sample-splitting); EL3 (the two-stage EL procedure); and IVX (the test of Kostakis et al. (2015) and Phillips and Lee (2016)). The sample period is from January 1952 to December 2024.

6 Empirical Application

This section is devoted to revisiting the evidence on the ability of financial and macroeconomic variables to predict stock market returns. Despite the voluminous literature on this subject, there is still a debate as to whether future stock returns are predictable or not. On one hand, studies like Lettau and Ludvigson (2001) argue that "...excess returns are predictable by variables such as dividend-price ratios, earnings-price ratios, dividend-earnings ratios, and an assortment of other financial indicators". On the other hand, however, studies like Welch and Goyal (2008) suggest that "...a healthy skepticism is appropriate when it comes to predicting the equity premium". We aim to shed some light on this debate by conducting a battery of mean and quantile predictability tests that we developed in this paper.

We collect monthly data on the following eleven variables that are commonly used in the literature as predictors of the aggregate market: the dividend payout ratio, the long-term yield, the dividend yield, the dividend-price ratio, the Treasury bill rate, the earnings-price ratio, the book-to-market value, the default yield spread, the net equity expansion, the term spread, and the inflation rate. The data is obtained from Amit Goyal's website[2] and covers the period from January 1952 to December 2024. The dependent variable in all predictive regressions is the continuously compounded return of the CRSP value weighted index[3] in excess of the one-month Treasury bill rate.

We first analyze empirical evidence of mean predictability. Table 5 reports the p-values for three testing procedures: the sample-splitting EL method in Section 2.2 (EL2), the two-stage

[2]See https://sites.google.com/view/agoyal145.

[3]The data for the CRSP index is obtained from Kenneth French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

Table 6. Empirical results for predictive quantile regressions

Quantile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Panel A: EL3									
Dividend price ratio	0.32	0.65	0.99	0.60	0.70	0.44	0.16	0.10	0.76
Dividend yield	0.12	0.44	0.82	0.74	0.77	0.48	0.21	0.14	0.62
Earnings price ratio	0.39	0.77	0.64	0.31	0.41	0.75	0.25	0.37	0.79
Dividend payout ratio	0.68	0.81	0.38	0.20	0.29	0.18	0.09	0.03	0.15
Book-to-market ratio	0.72	0.42	0.33	0.10	0.17	0.80	0.33	0.14	0.66
Net equity expansion	0.51	0.32	0.05	0.05	0.06	0.64	0.21	0.29	0.01
Treasury bill rate	0.20	0.00	0.00	0.00	0.01	0.04	0.70	0.84	0.54
Long term yield	0.44	0.02	0.00	0.00	0.06	0.05	0.74	0.54	0.76
The term spread	0.14	0.02	0.15	0.15	0.13	0.37	0.90	0.42	0.21
Default yield spread	0.11	0.11	0.41	0.93	0.40	0.09	0.00	0.00	0.00
Inflation	0.22	0.04	0.00	0.00	0.01	0.12	0.30	0.63	0.17
Panel B: IVX									
Dividend price ratio	0.00	0.00	0.01	0.89	0.97	0.28	0.04	0.00	0.00
Dividend yield	0.92	0.80	0.99	0.91	0.80	0.17	0.24	0.07	0.03
Earnings price ratio	0.85	0.16	0.04	0.29	0.71	0.65	0.39	0.37	0.10
Dividend payout ratio	0.04	0.23	0.44	0.02	0.07	0.08	0.19	0.01	0.04
Book-to-market ratio	0.00	0.00	0.01	0.15	0.73	0.04	0.00	0.08	0.00
Net equity expansion	0.87	0.28	0.29	0.33	0.39	0.97	0.41	0.35	0.07
Treasury bill rate	0.10	0.01	0.00	0.00	0.02	0.03	0.32	0.83	0.41
Long term yield	0.26	0.04	0.02	0.00	0.01	0.04	0.34	0.98	0.41
The term spread	0.06	0.11	0.46	0.78	0.61	0.66	0.77	0.70	0.86
Default yield spread	0.02	0.09	0.22	0.13	0.92	0.04	0.01	0.00	0.00
Inflation	0.18	0.15	0.01	0.00	0.00	0.04	0.25	0.30	0.12

Note: This table reports p-values for the null hypothesis of no predictability in conditional quantiles of excess stock returns. The dependent variable is the monthly continuously compounded return on the CRSP value-weighted index in excess of the one-month Treasury bill rate. Predictive variables are listed in the first column and results are shown across quantiles $\tau \in \{0.1, 0.2, \ldots, 0.9\}$. Tests are based on two procedures: EL3 (the two-stage EL procedure in Section 3.2) and IVX (the test of Lee (2016)). The sample period is January 1952 to December 2024.

EL approach proposed in Section 2.3 (EL3), and the IVX procedure of Kostakis et al. (2015) and Phillips and Lee (2016). The results indicate that the inflation rate and Treasury bill rate are significant predictors under all three tests at the 5% level. There is also some evidence of predictability for the dividend payout ratio and the long-term yield under IVX at the 10% level, although these variables are not significant under the EL-based tests. All other predictors are insignificant across methods, providing no statistical evidence to reject the null hypothesis of no predictability for those variables. Overall, these findings provide some evidence that inflation and short-term interest rates may have predictive content for market excess returns, while the evidence for other variables is weak and depends on the testing method. This pattern suggests that mean predictability, where present, is limited and concentrated in a few economically relevant predictors and highlights the consistency of EL-based tests with conventional methods.

We next examine empirical evidence of quantile predictability. Table 6 reports p-values for the two-stage EL-based test (EL3) and the IVX procedure across eleven predictors and quantiles $\tau \in \{0.1, 0.2, \ldots, 0.9\}$. Several interesting findings emerge. First, the Treasury bill rate and the long-term yield stand out as the most consistent predictors in the lower quantiles.

Both variables are highly significant at $0.2 \leq \tau \leq 0.6$ under EL3 and IVX, with p-values often below 1%. This suggests that short-term interest rates and bond yields contain predictive information for downside risk in equity returns. Inflation also shows strong significance in the lower tail ($0.2 \leq \tau \leq 0.5$), particularly under EL3, while its predictive content diminishes in the upper quantiles.

At central quantiles (say, $\tau = 0.50$), evidence of predictability weakens considerably across all variables. In the upper tail ($\tau \geq 0.70$), most predictors are insignificant, with the exception of the default yield spread, which becomes significant under both EL3 and IVX, indicating its relevance for extreme return outcomes. IVX occasionally flags dividend-related ratios and term spreads at the 10% level, but these signals are rather sporadic.

Overall, predictability appears concentrated in the lower tail of the return distribution, with interest rate variables and inflation emerging as the most relevant predictors, while the default yield spread dominates in the upper tail. Other predictors exhibit weak or inconsistent significance, reinforcing that predictability is limited to a small set of variables and primarily in the tails. This pattern shows the importance of modeling tail behavior and suggests that quantile-based tests can reveal forms of predictability that remain not captured in mean regressions.

7 Conclusion

This paper develops an empirical likelihood framework for unified inference in predictive mean and quantile regressions across a broad spectrum of persistence, including stationary, near-integrated, unit-root, and mildly explosive predictors. We establish Wilks-type chi-squared limits of the proposed test statistics uniformly across the persistence of the predictor. We handle the unknown intercepts through two complementary procedures: a sample-splitting approach that operates under relaxed regularity conditions, and a new two-stage method that uses the full sample to improve efficiency and extends naturally to quantile inference, where sample splitting is infeasible. We investigate higher-order finite-sample distortions under strong persistence, and propose gradually saturating weights and a Bartlett-type bias correction to mitigate the effects.

Simulation evidence demonstrates that the proposed EL procedures deliver well-calibrated size and competitive power across persistence regimes. In mean regressions, the two-stage method improves efficiency and sample splitting exhibits robust size properties even in highly persistent settings. In predictive quantile regressions, the EL approach performs particularly well, delivering close-to-nominal size across a wide range of quantiles including the tails, while retaining strong power relative to existing alternatives in many scenarios.

An empirical application to U.S. equity returns illustrates the practical value of the framework. Mean predictability is modest and concentrated in a small set of predictors such as inflation and short-term interest rates. Quantile-based inference uncovers richer heterogeneity

across the return distribution, with interest-rate variables and inflation showing stronger predictive content in the lower tail and the default yield spread becoming more informative in the upper tail. The results reveal that predictability is not uniform across the distribution, pointing to the importance of distributional analysis in asset returns predictability. The findings highlight the practical relevance of the proposed framework for risk management, portfolio allocation, and policy evaluation, offering a flexible and theoretically grounded tool for assessing predictability across the entire conditional distribution of returns.

References

Amihud, Y. and C. M. Hurvich (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis 39*(4), 813–841.

Andrews, D. and P. Guggenberger (2014). Asymptotics for ls, gls, and feasible gls statistics in an ar(1) model with conditional heteroskedasticity. *Journal of Econometrics 169*(2), 196–210.

Aue, A. and L. Horváth (2007). A limit theorem for mildly explosive autoregression with stable errors. *Econometric Theory 23*(2), 201–220.

Belloni, A., V. Chernozhukov, and D. Chetverikov (2019). Conditional quantile processes based on series or many regressors. *Journal of Econometrics 213*(1), 4–29.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes* (2nd ed.). Springer, Berlin.

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

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

Cai, Z., Y. Wang, and Y. Wang (2015). Testing instability in a predictive regression model with nonstationary regressors. *Econometric Theory 31*(5), 953–980.

Cai, Z. and Z. Xiao (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics 167*(2), 413–425.

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

Cavanagh, C. L., G. Elliott, and J. H. Stock (2009). Inference in models with nearly integrated regressors. *Econometric Theory 11*(5), 1131–1147.

Cenesizoglu, T. and A. Timmermann (2012). Do return prediction models add economic value? *Journal of Banking & Finance 36*(11), 2974–2987.

Choi, Y., S. Jacewitz, and J. Y. Park (2016). A reexamination of stock return predictability. *Journal of Econometrics 192*(1), 168–189.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106*(1), 157–181.

Demetrescu, M., I. Georgiev, M. M. Rodrigues, Paulo, and A. M. Taylor, Robert (2023). Extensions to IVX methods of inference for return predictability. *Journal of Econometrics 237*(1), 105–271.

Demetrescu, M. and P. M. M. Rodrigues (2022). Residual-augmented IVX predictive regression. *Journal of Econometrics 227*(2), 429–460.

DiCiccio, T., P. Hall, and J. Romano (1991). Empirical likelihood is bartlett-correctable. *Annals of Statistics 19*(2), 1053–1061.

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

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

Gao, J. and I. Gijbels (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association 103*, 1584–1594.

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

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *Review of Financial Studies 33*(5), 2223–2273.

Hall, P. and C. Heyde (1980). The central limit theorem. In P. Hall and C. Heyde (Eds.), *Martingale Limit Theory and its Application*, pp. 51–96. Canberra: Academic Press.

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

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics 12*(4), 371–389.

Kitamura, Y., G. Tripathi, and H. Ahn (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica 72*(6), 1667–1714.

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

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

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

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

Lee, J. H. (2018). Limit theory for explosive autoregression under conditional heteroskedasticity. *Journal of Statistical Planning and Inference 196*(1), 30–55.

Lettau, M. and S. Ludvigson (2001). Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy 109*(6), 1238–1287.

Li, C., D. Li, and L. Peng (2017). Uniform test for predictive regression with ar errors. *Journal of Business & Economic Statistics 35*(1), 29–39.

Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society, Series B 67*(3), 381–393.

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

Liu, Y. and J. Chen (2010). Adjusted empirical likelihood with high-order precision. *Annals of Statistics 38*(3), 1341–1362.

Magdalinos, T. and P. C. B. Phillips (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory 25*(2), 482–526.

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

Otsu, T. (2008). Conditional empirical likelihood estimation and inference for quantile regression models. *Journal of Econometrics 142*(2), 508–538.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*(2), 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics 18*(1), 90–120.

Park, J. Y. (2002). Nonstationary nonlinear heteroskedasticity. *Journal of Econometrics 110*(2), 383–415.

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

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica 55*, 277–301.

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

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

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

Phillips, P. C. B. and T. Magdalinos (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics 136*(1), 115–130.

Phillips, P. C. B., Y. Wu, and J. Yu (2011). Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International Economic Review 52*(1), 201–226.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

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

Wang, H. J. and Z. Zhu (2011). Empirical likelihood for quantile regression models with longitudinal data. *Journal of Statistical Planning and Inference 141*(4), 1603–1615.

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

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

Yang, B., X. Liu, L. Peng, and Z. Cai (2021). Unified tests for a dynamic predictive regression. *Journal of Business & Economic Statistics 39*(3), 684–699.

Yang, B., W. Long, L. Peng, and Z. Cai (2020). Testing the predictability of u.s. housing price index returns based on an IVX-AR model. *Journal of the American Statistical Association 115*(532), 1598–1619.

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

Appendix: Proofs of the Main Results

We use C and C' to denote some generic constants whose value may vary between occurrences, unless defined otherwise. Without loss of generality, we assume $X_0 = 0$ following standard practice. The norm $\|.\|_1$ is taken to mean the L_1-norm.

Proof of Theorem 1. The proof is a special case of the proof of Theorem 3 below, with α being a fixed, known constant, and hence is omitted for brevity. □

Proof of Theorem 2. We establish the asymptotic normality via the martingale central limit theorem, Corollary 3.1 of Hall and Heyde (1980). First, the effect of θ in equation (2) can be eliminated by employing similar argument in the proof of Theorem 2 in Zhu et al. (2014), although their error terms are assumed to be strictly i.i.d. The next step concerns the derivation of the probability limit of the martingale conditional variance. We write $\mathcal{H}_{m,s} = \sigma(U_s^*, U_{s-1}^*, \ldots, U_0^*, \varepsilon_s^*, \varepsilon_{s-1}^* \ldots, \varepsilon_0^*)$ with $U_s^* = U_{s+m} - U_s$, $\varepsilon_s^* = \varepsilon_{s+m} - \varepsilon_s$, and recall that U_s is a martingale difference with respect to $\mathcal{G}_{s-1} = \sigma(\{V_r, z_r\} : r \leq s-1)$.

Define $Z_{Tt}^* := (Y_t^* - \beta X_{t-1}^*)w(X_{t-1}^*) = U_t^* w(X_{t-1}^*)$. If we write $\mu_t := \mathbb{E}(Z_{Tt}^*|\mathcal{H}_{m,t-1})$ and $\check{Z}_{Tt}^* := Z_{Tt}^* - \mu_t$, then $\{\check{Z}_{Tt}^*\}$ is a martingale difference array with respect to $\mathcal{H}_{m,t-1}$. We later show in (25) that $m^{-1/2}\sum_{t=1}^m \mu_t = o_p(1)$ and $\frac{1}{m}\sum_t \mathbb{E}(\mu_t^2) \leq Cm^{-1}\sum_k \alpha(k)^c \to 0$. Under conditional homoscedasticity of U_s (i.e. Assumption 2(i)), we have from (8) that

$$
\begin{aligned}
& \frac{1}{m}\sum_{t=1}^m \mathbb{E}\big(\check{Z}_{Tt}^{*,2}|\mathcal{H}_{m,t-1}\big) \;=\; \frac{1}{m}\sum_{t=1}^m \mathbb{E}\big(Z_{Tt}^{*,2}|\mathcal{H}_{m,t-1}\big) + o_p(1) \\
= \;& \frac{1}{m}\sum_{t=1}^m \mathbb{E}\left(\Big[(Y_t^* - \beta X_{t-1}^*)w(X_{t-1}^*)\Big]^2 \Big| \mathcal{H}_{m,t-1}\right) + o_p(1) \\
= \;& \frac{1}{m}\sum_{t=1}^m \mathbb{E}\left(U_t^{*,2} w(X_{t-1}^*)^2 \Big| \mathcal{H}_{m,t-1}\right) + o_p(1) \\
= \;& \frac{1}{m}\sum_{t=1}^m \big[\mathbb{E}(U_{t+m}^2|\mathcal{H}_{m,t-1}) \;+\; \mathbb{E}(U_t^2|\mathcal{H}_{m,t-1}) + 0\big]\cdot w(X_{t-1}^*)^2 + o_p(1) \\
= \;& \frac{1}{m}\sum_{t=1}^m \big[\sigma_U^2 \;+\; \mathbb{E}\left(\mathbb{E}(U_t^2|\mathcal{G}_{t-1+m})\big|\mathcal{H}_{m,t-1}\right)\big]\cdot w(X_{t-1}^*)^2 + o_p(1) \\
= \;& \sigma_U^2 \frac{1}{m}\sum_{t=1}^m w(X_{t-1}^*)^2 \;+\; \frac{1}{m}\sum_{t=1}^m g_t w(X_{t-1}^*)^2 + o_p(1) \\
= \;& \frac{2\sigma_U^2}{m}\sum_{t=1}^m w(X_{t-1}^*)^2 + \frac{1}{m}\sum_{t=1}^m (g_t - \sigma_U^2) w(X_{t-1}^*)^2 + o_p(1) \;=\; \mathcal{Q}_{m,1} + \mathcal{Q}_{m,2} + o_p(1) \qquad (19)
\end{aligned}
$$

by the tower property of conditional expectations.

Now, write $\mathcal{B} := \sigma(\{V_{j+m}, z_{j+m}\} : 0 \leq j \leq t-1, j+m \geq t)$ so that upon noting $j \in \{0, 1, \ldots, t-1\}$ we see that $\mathcal{B} = \sigma(V_m, z_m, V_{m+1}, z_{m+1}, \ldots, V_{t-1+m}, z_{t-1+m})$. It is then straightforward to show that $(\mathcal{G}_{t-1} \vee \mathcal{H}_{m,t-1}) \subset (\mathcal{G}_{t-1} \vee \mathcal{B})$, where $\mathcal{C} \vee \mathcal{D} = \sigma(\mathcal{C} \cup \mathcal{D})$, the join of σ-fields.

Therefore, we have

$$\begin{aligned}
\|g_t - \sigma_U^2\|_1 &= \|\mathbb{E}(U_t^2|\mathcal{H}_{m,t-1}) - \mathbb{E}(U_t^2|\mathcal{G}_{t-1})\|_1 \\
&\leq \|\mathbb{E}(U_t^2|\mathcal{G}_{t-1} \vee \mathcal{H}_{m,t-1}) - \mathbb{E}(U_t^2|\mathcal{G}_{t-1})\|_1 \\
&\leq \|\mathbb{E}(U_t^2|\mathcal{G}_{t-1} \vee \mathcal{B}) - \mathbb{E}(U_t^2|\mathcal{G}_{t-1})\|_1 \\
&= \|\mathbb{E}[U_t^2 - \mathbb{E}(U_t^2|\mathcal{G}_{t-1})|\mathcal{G}_{t-1} \vee \mathcal{B}]\| \\
&\leq \sup_{\substack{Y \in L^\infty(\mathcal{G}_{t-1} \vee \mathcal{B}) \\ \|Y\|_\infty \leq 1}} \left|\mathbb{E}[(U_t^2 - \mathbb{E}(U_t^2|\mathcal{G}_{t-1}))Y]\right| \\
&\leq C \sup_{\substack{Y \in L^\infty(\mathcal{B}) \\ \|Y\|_\infty \leq 1}} \left|\mathrm{Cov}(\zeta_t, Y)\right|, \qquad (20)
\end{aligned}$$

because $\mathbb{E}(U^2|\mathcal{F}) - \mathbb{E}(U^2|\mathcal{A}) = \mathbb{E}[U^2 - \mathbb{E}(U^2|\mathcal{A})|\mathcal{F}]$ for $\mathcal{A} \subset \mathcal{F}$, and for integrable $W := U_t^2 - \mathbb{E}(U_t^2|\mathcal{G}_{t-1})$ and $\|Y\|_\infty \leq 1$, we have (i) $\mathbb{E}[WY] = \mathbb{E}[\mathbb{E}(W|\mathcal{F})Y] \leq \mathbb{E}[\mathbb{E}(W|\mathcal{F})|Y|] \leq \|\mathbb{E}(W|\mathcal{F})\|_1$, and (ii) choosing $Y^* := \mathrm{sgn}(\mathbb{E}(W|\mathcal{F}))$ we have $Y^* \in L^\infty(\mathcal{F})$, $\|Y^*\|_\infty \leq 1$, and $\sup_Y |\mathbb{E}[WY]| \geq \mathbb{E}[WY^*] = \mathbb{E}[\mathbb{E}(W|\mathcal{F})\mathrm{sgn}(\mathbb{E}(W|\mathcal{F}))] = \mathbb{E}|\mathbb{E}(W|\mathcal{F})| \equiv \|\mathbb{E}(W|\mathcal{F})\|_1$. Here $\zeta_t = U_t^2 - \mathbb{E}(U_t^2|\mathcal{G}_{t-1})$.

Since $\sigma(\zeta) \subset \sigma((V_s, z_s) : s \leq t)$, $\sigma(Y) \subset \mathcal{B}^+ = \sigma((V_s, z_s) : s \geq m)$, and $\sup_t \mathbb{E}|U_t|^{2+q} < \infty$, for $p = (2+q)/q$ we have

$$\begin{aligned}
|\mathrm{Cov}(\zeta_t, Y)| &\leq C\alpha(m-t)^{1/p}\|\zeta_t\|_{(2+q)/2}\|Y\|_\infty \\
&\leq C'\alpha(m-t)^{1/p}\|U_t^2\|_{(2+q)/2} \qquad (21)
\end{aligned}$$

for each $t = 1, \ldots, m$ by Davydov's inequality (see, for example, Corollary 1.1 in Bosq (1998)).

Note that for all weight function choices we consider: $w(x) = x/\sqrt{1+x^2}$, $w(x) = x/(1+|x|)$, and $w(x) = \tanh(x/b)$, we have $w(X_{t-1})^2 = O_p(1)$. Hence, in view of (20) and (21), the Cesàro mean theorem yields

$$\begin{aligned}
\|\mathcal{Q}_{m,2}\|_1 &= \left\| \frac{1}{m} \sum_{t=1}^m (g_t - \sigma_U^2) w(X_{t-1}^*)^2 \right\|_1 \\
&\leq \frac{1}{m} \sum_{t=1}^m \left\| g_t - \sigma_U^2 \right\|_1 \leq \frac{C}{m} \sum_{t=1}^m \alpha(m-t)^{1/p} \\
&\leq \frac{1}{m} \sum_{k=0}^{m-1} \alpha(k)^{1/p} \to 0, \qquad (22)
\end{aligned}$$

because $\alpha(k) \to 0$ as $k \to \infty$, and $0 \leq \alpha(k) \leq 1$.

In the meantime, since $\mu_t = w(X_{t-1}^*)\mathbb{E}(U_{t+m} - U_t|\mathcal{H}_{m,t-1}) = -w(X_{t-1}^*)\mathbb{E}(U_t|\mathcal{H}_{m,t-1})$ and $\mathcal{H}_{m,t-1} \subset (\mathcal{G}_{t-1} \vee \mathcal{B}_t)$, using similar idea as in (20), (21), and Davydov's inequality, we have

$$\begin{aligned}
\|\mathbb{E}(U_t|\mathcal{H}_{m,t-1})\|_1 &\leq \|\mathbb{E}(U_t|\mathcal{G}_{t-1} \vee \mathcal{B}_t)\|_1 \\
&\leq C \sup_{\substack{Y \in L^\infty(\mathcal{B}_t) \\ \|Y\|_\infty \leq 1}} \left|\mathrm{Cov}(U_t, Y)\right| \qquad (23) \\
&\leq C'\alpha(m-t)^\delta \|U_t\|_{2+q}\|Y\|_\infty \qquad (24)
\end{aligned}$$

where $\delta = 1 - 1/(2+q) = (1+q)/(2+q)$.

Since by Assumption 1, the mixing rate is either geometric or polynomial with exponent greater than $2 + 1/(2(1+q))$, we have

$$\begin{aligned}\left\| \frac{1}{\sqrt{m}} \sum_{t=1}^{m} \mu_t \right\|_1 &\leq \frac{1}{\sqrt{m}} \sum_{t=1}^{m} \|\mu_t\|_1 \\ &\leq \frac{C}{\sqrt{m}} \sum_{t=1}^{m} \alpha(m-t)^{\frac{1+q}{2+q}} = \frac{C}{\sqrt{m}} \sum_{k=0}^{m-1} \alpha(k)^{\frac{1+q}{2+q}} \to 0. \end{aligned} \tag{25}$$

Furthermore, following the same argument and using the L_2-norm instead, we can also immediately see that $\frac{1}{m}\sum_t \mathbb{E}(\mu_t^2) \leq Cm^{-1}\sum_k \alpha(k)^c \to 0$.

Returning to (22), since $\mathcal{Q}_{m,2} = o_p(1)$, we finally have

$$\frac{1}{m}\sum_{t=1}^{m} \mathbb{E}\big(\check{Z}_{Tt}^{*,2} | \mathcal{H}_{m,t-1}\big) = 2\sigma_U^2 \cdot \left\{ \frac{1}{m}\sum_{t=1}^{m} w(X_{t-1}^*)^2 \right\} + o_p(1). \tag{26}$$

Now, we note that in the mildly integrated case, i.e., $\rho = \rho_T = 1 + c/T^a$ with $0 < a < 1$ and $c < 0$, Phillips and Magdalinos (2007) showed that

$$T^{-a/2} X_{\lfloor T^a t \rfloor} \implies \int_0^t e^{c(t-r)} dW(r), \tag{27}$$

where W is Brownian motion with variance $\sigma^2 = \mathbb{E}(\varepsilon_t^2)$ and $\Longrightarrow$ refers to weak convergence in the Skorohod space $\mathcal{D}[0,\ell]$ (i.e. the space of the collection of $\mathbb{R}$-valued càdlàg functions on $[0,1]$), see e.g. Pollard (1984). The initial condition $X_0 = o_p(T^{a/2})$ is imposed, and a finite moment strictly higher than 2 is required for the i.i.d. error term, which is consistent with what we assume.

Furthermore, in the mildly explosive case, i.e., $\rho = \rho_T = 1 + c/T^a$ with $0 < a < 1$ and $c > 0$, the proof of Aue and Horváth (2007) suggests that for any fixed constant $\ell > 0$ we have

$$\frac{1}{\xi_T^{-1/2}(\mathbb{E}(\varepsilon_1^2))^{1/2}} \rho^{-\lfloor \ell/\xi_T \rfloor} X_{\lfloor \ell/\xi_T \rfloor} \implies e^{-\ell}\mathcal{W}_{\alpha_h,\beta_h}(\ell) + \int_0^\ell \mathcal{W}_{\alpha_h,\beta_h}(x)dx, \tag{28}$$

where $\xi_T = \log\rho = \log\rho_T = \log(1 + c/T^a) \to 0$ as $T \to \infty$, and $\mathcal{W}_{\alpha_h,\beta_h}$ is a strictly α-stable random variable.

In the near integrated case where $a = 1$ and $c \neq 0$, we know from Phillips (1987) that

$$\frac{1}{\sqrt{T}} X_{\lfloor Tr \rfloor} \implies \int_0^r e^{-c(r-s)} dW(s). \tag{29}$$

In all three cases, the denominators of the "multiplier" to X all tend to the infinity in the LHS of (27), (28) and (29). Specifically, we have $|X_t^*| \xrightarrow{p} +\infty$ (i.e. $\mathbb{P}(|X_t^*| > r) \to 1$ for every $r > 0$.

By Skorokhod representation theorem and Lebesgue's dominated convergence theorem, we have

$$w(X_{t-1}^*)^2 \xrightarrow{L_1} 1 \tag{30}$$

as $t \to \infty$.

For example, we have

$$\frac{(X_{t-1}^*)^2}{1+(X_{t-1}^*)^2} \xrightarrow{L_1} 1;\ \frac{(X_{t-1}^*)^2}{(1+|X_{t-1}^*|)^2} \xrightarrow{L_1} 1;\ \tanh^2(X_{t-1}^*/b) \xrightarrow{L_1} 1 \tag{31}$$

as $t \to \infty$, for any fixed $b > 0$. Consequently, the stochastic convergence of Cesàro means of random variables yields

$$\frac{1}{m}\sum_{t=1}^{m} w(X_{t-1}^*)^2 \xrightarrow{L_1} 1 \tag{32}$$

as $T \to \infty$, which implies convergence in probability to 1. Therefore, in view of (19), (26), and (32) we finally have

$$\frac{1}{m}\sum_{t=1}^{m} \mathbb{E}\big(\check{Z}_{Tt}^{*,2} | \mathcal{H}_{m,t-1}\big) \xrightarrow{p} 2\sigma_U^2 \tag{33}$$

in all nonstationary cases we consider.

Meanwhile, in the stationary case where $|\rho| < 1$, since $w(\cdot)$ is bounded and continuous (hence measurable), we have

$$\mathbb{E}\left[w\left(X_{t-1}^*\right)^2\right] = \mathbb{E}\left\{w\left(\sum_{i=1}^{t-1}\rho^{t-i-1}\varepsilon_i^*\right)^2\right\} + o(1).$$

We write $\lim_{t\to\infty}\mathbb{E}\{\cdot\} =: \nu^2$ so that $\mathbb{E}\{w(X_{t-1}^*)^2\} = \nu^2 + o(1)$ as $t \to \infty$. Here the linear process $\sum_{k=0}^{\infty}\rho^k\varepsilon_{t-1-k}^*$ is well defined and converges absolutely a.s. since $\sum_{k\geq 0}|\rho|^k < \infty$; moreover w^2 is bounded, so dominated convergence yields the limit. For example, when $w(x) = x/\sqrt{1+x^2}$,

$$\begin{aligned}
\mathbb{E}\left[\frac{(X_{t-1}^*)^2}{1+(X_{t-1}^*)^2}\right] &= \mathbb{E}\left[\frac{\left(\sum_{i=1}^{t-1}\rho^{t-i-1}\varepsilon_i^*\right)^2}{1+\left(\sum_{i=1}^{t-1}\rho^{t-i-1}\varepsilon_i^*\right)^2}\right] + o_p(1) \\
&\to \lim_{t\to\infty}\mathbb{E}\left\{\frac{\left(\sum_{i=1}^{t-1}\rho^{t-i-1}\varepsilon_i^*\right)^2}{1+\left(\sum_{i=1}^{t-1}\rho^{t-i-1}\varepsilon_i^*\right)^2}\right\} + o_p(1) \\
&=: \nu^2 + o_p(1)
\end{aligned} \tag{34}$$

because the series converges absolutely almost surely.

Therefore, by the law of large numbers for stationary processes, as $m = \lfloor T/2 \rfloor \to \infty$ it follows that

$$\frac{1}{m}\sum_{t=1}^{m} \mathbb{E}\big(\check{Z}_{Tt}^{*,2} | \mathcal{H}_{m,t-1}\big) \xrightarrow{p} 2\sigma_U^2 \cdot \nu^2. \tag{35}$$

By Assumption 1, Markov's inequality, and Jensen's inequality, for any $\epsilon > 0$ we have

$$\begin{aligned}
&\sum_{t=1}^{m} \mathbb{E}\left(\frac{\check{Z}_{Tt}^{*,2}(\beta)}{m} 1\left\{ \frac{\check{Z}_{Tt}^{*,2}(\beta)}{m} > \epsilon \right\} \middle| \mathcal{H}_{m,t-1} \right) \\
&\quad \leq \sum_{t=1}^{m} \frac{1}{\epsilon^{q/2}} \mathbb{E}\left(\left| \frac{1}{\sqrt{m}} \check{Z}_{Tt}^{*}(\beta) \right|^{2+q} \middle| \mathcal{H}_{m,t-1} \right) \\
&\quad \leq \frac{C}{\epsilon^{q/2} m^{1+q/2}} \sum_{t=1}^{m} \mathbb{E}\left(|U_t^*|^{2+q} \left| w(X_{t-1}^*) \right|^{2+q} \middle| \mathcal{H}_{m,t-1} \right) \\
&\quad \leq \frac{\sup_t \mathbb{E}|U_t^*|^{2+q}}{\epsilon^{q/2} m^{q/2}} \cdot \left\{ \frac{1}{m} \sum_{t=1}^{m} \left| w(X_{t-1}^*) \right|^{2+q} \right\} \;=\; O_p(1) \cdot \frac{1}{m^{q/2}} = o_p(1). \qquad (36)
\end{aligned}$$

Now that the conditional Lindeberg condition (36) is met, it follows by the martingale central limit theorem (e.g. Hall and Heyde (1980)) that

$$\frac{1}{\sqrt{m}} \sum_{t=1}^{m} \check{Z}_{Tt}^{*}(\beta) \xrightarrow{d} N\big(0, \eta^2\big), \qquad (37)$$

where η^2 is the probability limit of (19), i.e. $2\sigma_U^2 \cdot \nu^2$ in the stationary case and $2\sigma_U^2$ otherwise.

The same result holds under conditional heteroscedasticity of U_t, i.e. Assumption 2-(ii) as we show now. Consider

$$\frac{1}{m} \sum_{t=1}^{m} \left[\mathbb{E}(U_{t+m}^2 | \mathcal{H}_{m,t-1}) + \mathbb{E}(U_t^2 | \mathcal{H}_{m,t-1}) \right] w(X_{t-1}^*)^2 = \mathcal{W}_{1,t} + \mathcal{W}_{2,t}.$$

For the first term, using the same arguments used before and the Cesàro mean theorem we have

$$\begin{aligned}
\left\| \mathbb{E}(U_{t+m}^2 | \mathcal{H}_{m,t-1}) - \sigma_{t+m}^2 \right\|_1 &= \left\| \mathbb{E}\big(\mathbb{E}[U_{t+m}^2 | \mathcal{G}_{t-1+m}] | \mathcal{H}_{m,t-1} \big) - \sigma_{t+m}^2 \right\|_1 \\
&= \left\| \mathbb{E}\big(\sigma_{t+m}^2 | \mathcal{H}_{m,t-1} \big) - \sigma_{t+m}^2 \right\|_1 \\
&\leq C\alpha(m-t)^{q/(2+q)} \to 0. \qquad (38)
\end{aligned}$$

Similarly, as for the second term, we have

$$\left\| \mathbb{E}(U_t^2 | \mathcal{H}_{m,t-1}) - \mathbb{E}(U_t^2 | \mathcal{G}_{t-1}) \right\|_1 \leq C\alpha(m-t)^{q/(2+q)} \to 0. \qquad (39)$$

Hence, it follows by the triangle inequality that

$$\mathcal{W}_{1,t} + \mathcal{W}_{2,t} \;=\; \frac{1}{m} \sum_{t=1}^{m} (\sigma_{t+m}^2 + \sigma_t^2) w(X_{t-1}^*)^2 \;+\; o_p(1), \qquad (40)$$

because $0 \leq w(X_{t-1}^*)^2 \leq 1$, a.s., $w(X_{t-1}^*)^2$ converges in probability to a constant as $t \to \infty$ for all persistence classes C1−C5, and $\sup_t \mathbb{E}(\sigma_t^{2+q/2}) < \infty$.

Furthermore, with the uniform boundedness of $\mathbb{E}\sigma_t^2$ we see that

$$\frac{1}{m}\sum_{t=1}^{m}(\sigma_t^2+\sigma_{t+m}^2) \;=\; \frac{1}{m}\left(\sum_{s=1}^{m}\sigma_s^2+\sum_{s=m+1}^{2m}\sigma_s^2\right) \;=\; 2\frac{1}{2m}\sum_{s=1}^{2m}\sigma_s^2 \tag{41}$$

and therefore, in view of (40), (41) and Assumption 2, it follows that $\mathcal{W}_{1,t}+\mathcal{W}_{2,t}$ converges in probability to $2\sigma^2\cdot\nu^2$ in the stationary case and $2\sigma^2$ otherwise. The conditional Lindeberg condition can be verified by following the same argument as above, using the boundedness of the weight and $\mathbb{E}|U_t|^{2+q}$, and is not repeated for brevity.

We now check if X converges in distribution to a random variable (that is finite a.s.) in the conditionally heteroscedastic case. In the mildly explosive case, by Lemma 4.1 of Lee (2018) we have

$$\frac{1}{T^{a/2}\rho_T^T}X_T \;\Longrightarrow\; X_c \tag{42}$$

where X_c is centred Gaussian random variable with variance $\mathbb{E}(\sigma_t^2)/(2c)$. The cases of unit root and mildly integrated regressors can be handled based on the results by Andrews and Guggenberger (2014).

As for the near integrated case, Lemma 3.1 of Lee (2018) showed that the same limit theory of Phillips (1987) in the conditional homoscedastic case is valid: i.e.,

$$\frac{1}{\sqrt{T}}X_{\lfloor Tr\rfloor} \;\Longrightarrow\; \int_0^r e^{c(r-s)}dW(s). \tag{43}$$

Consequently, we see that the previous argument continues to apply, and

$$w(X_{t-1}^*)^2 \;\xrightarrow{L_1}\; 1. \tag{44}$$

For example, for $w(x)=x/\sqrt{1+x^2}$ we have $(X_{t-1}^*)^2/(1+(X_{t-1}^*)^2)\to^p 1$.

Therefore, as before in (37) it follows that

$$\frac{1}{\sqrt{m}}\sum_{t=1}^{m}\check{Z}_{Tt}^*(\beta) \;\xrightarrow{d}\; N\big(0,\eta^2\big), \tag{45}$$

where η^2 is equal to $2\sigma^2\cdot\nu^2$ in the stationary case and $2\sigma^2$ otherwise.

Finally, following the same derivations in (16) we have

$$\ell_T^*(\beta) \;=\; \frac{\left(\frac{1}{\sqrt{m}}\sum_{t=1}^{m}\check{Z}_t^*(\beta)\right)^2}{\frac{1}{m}\sum_{t=1}^{m}(\check{Z}_t^*(\beta))^2}+o_p(1) \;\xrightarrow{d}\; \chi_1^2, \tag{46}$$

which completes the proof. □

Proof of Theorem 3. With $\widetilde{Y}_t = Y_t - \widetilde{\alpha}$ and $\widetilde{Z}_t(\beta) = \big[\widetilde{Y}_t - \beta X_{t-1}\big] \cdot w^c(X_{t-1})$, recall that the centered weight is

$$w^c(X_{t-1}) := w(X_{t-1}) - \frac{1}{T}\sum_{s=1}^{T} w(X_{s-1}). \tag{47}$$

We have

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{Z}_t(\beta) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\big[\widetilde{Y}_t - \beta X_{t-1}\big]\cdot w^c(X_{t-1}) \\
= &\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big[(Y_t - \alpha - \beta X_{t-1}) - (\widetilde{\alpha} - \alpha)\Big]\cdot w^c(X_{t-1}) \\
= &\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_t w^c(X_{t-1}) - \sqrt{T}(\widetilde{\alpha} - \alpha)\frac{1}{T}\sum_{t=1}^{T} w^c(X_{t-1}) = A_T - B_T,
\end{aligned}$$

and B_T is zero by the construction of the centered weight.

In order to employ the martingale central limit theorem, it remains to establish (i) an asymptotic limit for A_T, and (ii) consistency of the average $T^{-1}\sum_{t=1}^{T}(U_t w^c(X_{t-1}))^2$, as well as checking the conditional Lindeberg condition. For simplicity of presentation we write

$$w_t := w(X_{t-1}), \qquad \bar{w}_T := \frac{1}{T}\sum_{s=1}^{T} w_s, \qquad w_t^c := w_t - \bar{w}_T, \tag{48}$$

so that $A_T = T^{-1/2}\sum_{t=1}^{T} U_t w_t^c$. Define also

$$A_{1T} := \frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_t w_t, \qquad A_{0T} := \frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_t, \tag{49}$$

so that $A_T = A_{1T} - \bar{w}_T A_{0T}$. Given $\mathcal{G}_t := \sigma(\{V_s, z_s\} : s \le t)$, since U_t is a martingale difference with respect to $\mathcal{G}_{t-1}$ and $w_t = w(X_{t-1})$ is $\mathcal{G}_{t-1}$-measurable, the vector

$$\Delta_{Tt} := \begin{pmatrix} U_t w_t \\ U_t \end{pmatrix} \tag{50}$$

is a martingale difference array with respect to $\mathcal{G}_{t-1}$. Consider the normalized partial sums

$$M_T := \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta_{Tt} = \begin{pmatrix} A_{1T} \\ A_{0T} \end{pmatrix}. \tag{51}$$

Write $\sigma_t^2 := \mathbb{E}(U_t^2 \mid \mathcal{G}_{t-1})$, which equals σ_U^2 under Assumption 2(i) and may be time-varying under Assumption 2(ii). The predictable quadratic variation matrix of M_T is

$$\langle M\rangle_T := \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big(\Delta_{Tt}\Delta_{Tt}'|\mathcal{G}_{t-1}\big) = \frac{1}{T}\sum_{t=1}^{T}\sigma_t^2\begin{pmatrix} w_t^2 & w_t \\ w_t & 1 \end{pmatrix}. \tag{52}$$

The weight assumptions (i.e. boundedness and asymptotic stability of Cesàro averages of w_t^2 across persistence classes $C1 - C5$) imply that $\langle M \rangle_T$ converges in probability to a finite, possibly random, limit matrix Σ. Consequently, the scalar quadratic variation associated with the centered combination satisfies

$$\eta_T^2 := \begin{pmatrix} 1 & -\bar{w}_T \end{pmatrix} \langle M \rangle_T \begin{pmatrix} 1 \\ -\bar{w}_T \end{pmatrix} = \frac{1}{T} \sum_{t=1}^{T} \sigma_t^2 (w_t^c)^2 \xrightarrow{p} \eta^2, \tag{53}$$

where η^2 vary across persistence classes. We emphasize that unlike in Theorem 2, η^2 here depends on the probability limit of $\bar{w}_T$ (instead of w_t^2), and hence may be random.

Next, we check the conditional Lindeberg condition. Since the weight satisfies $|w(\cdot)| \leq 1$, a.s., we have $|w_t^c| \leq |w_t| + |\bar{w}_T| \leq 2$, hence $\|\Delta_{Tt}\| \leq C|U_t|$. Using $\sup_t \mathbb{E}|U_t|^{2+q} < \infty$ from Assumption 1, for any $\varepsilon > 0$, it follows that

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\Big(\|\Delta_{Tt}\|^2 \mathbf{1}\{\|\Delta_{Tt}\| > \varepsilon \sqrt{T}\} \Big| \mathcal{G}_{t-1} \Big) \leq \frac{C}{T^{q/2}} \cdot \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\big(|U_t|^{2+q} | \mathcal{G}_{t-1} \big) = o_p(1).$$

Therefore, by the martingale central limit theorem we obtain the following stable convergence to mixed-normal:

$$M_T \xrightarrow{st} MN(0, \Sigma), \tag{54}$$

which implies

$$A_T = \begin{pmatrix} 1 & -\bar{w}_T \end{pmatrix} M_T \xrightarrow{st} MN(0, \eta^2). \tag{55}$$

In particular, when η^2 is nonrandom, depending on the persistence class, this reduces to the usual $N(0, \eta^2)$ limit. Consequently, we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_t w^c(X_{t-1}) \xrightarrow{st} MN(0, \eta^2). \tag{56}$$

Write $V_T := T^{-1} \sum_{t=1}^{T} \big(U_t w_t^c \big)^2$. We have

$$V_T = \frac{1}{T} \sum_{t=1}^{T} U_t^2 (w_t - \bar{w}_T)^2 = \frac{1}{T} \sum_{t=1}^{T} U_t^2 w_t^2 - 2\bar{w}_T \frac{1}{T} \sum_{t=1}^{T} U_t^2 w_t + \bar{w}_T^2 \frac{1}{T} \sum_{t=1}^{T} U_t^2 = A_T^{(2)} + B_T^{(2)} + C_T^{(2)}$$

Similarly, we define

$$A_T^{(2,0)} := \frac{1}{T} \sum_{t=1}^{T} \sigma_t^2 w_t^2, \quad B_T^{(2,0)} := \frac{1}{T} \sum_{t=1}^{T} \sigma_t^2 w_t, \quad C_T^{(2,0)} := \frac{1}{T} \sum_{t=1}^{T} \sigma_t^2. \tag{57}$$

Since w_t is $\mathcal{G}_{t-1}$-measurable, for $k = 0, 1, 2$, each difference $U_t^2 w_t^k - \mathbb{E}(U_t^2 \mid \mathcal{G}_{t-1}) w_t^k = \big(U_t^2 - \sigma_t^2 \big) w_t^k$ is a martingale difference.

With $\sup_t \mathbb{E}|U_t|^{2+q} < \infty$ and bounded w_t, the standard martingale law of large numbers

yields $A_T^{(2)} - A_T^{(2,0)} = o_p(1)$, $B_T^{(2)} - B_T^{(2,0)} = o_p(1)$, and $C_T^{(2)} - C_T^{(2,0)} = o_p(1)$. Therefore, we have

$$\begin{aligned} V_T =& \; A_T^{(2,0)} - 2\bar{w}_T B_T^{(2,0)} + \bar{w}_T^2 C_T^{(2,0)} + o_p(1) \\ =& \; \frac{1}{T}\sum_{t=1}^{T} \sigma_t^2 (w_t^c)^2 + o_p(1) \;=\; \eta_T^2 + o_p(1) \;\xrightarrow{\;p\;}\; \eta^2. \end{aligned} \tag{58}$$

Now, in view of (56) and (58), we have

$$\frac{A_T}{\sqrt{V_T}} \xrightarrow{\;st\;} N(0,1), \qquad \text{and hence} \qquad \frac{A_T^2}{V_T} \xrightarrow{\;st\;} \chi_1^2. \tag{59}$$

Hence, standard Taylor expansion argument as in equations (15) as before yields $\widetilde{\ell}_T(\beta) \to^{st} \chi_1^2$ as $T \to \infty$, which implies the desired convergence in distribution. □

Proof of Theorem 4. The proof is a special case of the proof of Theorem 5 below, with α_τ being a fixed, known constant, and hence is omitted for brevity. □

Proof of Theorem 5. We write $\psi_\tau(u) = \tau - 1(u < 0)$, $w_t := w(X_{t-1})$, $\overline{w}_T := T^{-1}\sum_{s=1}^{T} w_s$, and $w_t^c := w_t - \bar{w}_T$. With the two-stage intercept-adjusted response $\widetilde{Y}_t := Y_t - \widetilde{\alpha}_\tau$, we have

$$\widetilde{\xi}_{t,\tau}(\beta_\tau) \;:=\; \psi_\tau(\widetilde{Y}_t - \beta_\tau X_{t-1}) w_t^c \;=\; \psi_\tau(Y_t - \widetilde{\alpha}_\tau - \beta X_{t-1}) w_t^c. \tag{60}$$

Since there exists a constant $f_\tau(0) \in (0,\infty)$ and $\varepsilon_0 > 0$ such that, uniformly in t, $F_{t,\tau}(u|\mathcal{G}_{t-1}) := \mathbb{P}(U_{t,\tau} \le u|\mathcal{G}_{t-1}) = \tau + f_\tau(0)\, u + O(u^2)$ a.s. for $|u| \le \varepsilon_0$ by Assumption 3, we have

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \widetilde{\xi}_{t,\tau}(\beta_\tau) \;&=\; \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \psi_\tau(U_{t,\tau}) w_t^c + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Big\{ \psi_\tau(U_{t,\tau} - (\widetilde{\alpha}_\tau - \alpha_\tau)) - \psi_\tau(U_{t,\tau}) \Big\} w_t^c \\ &=:\; \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \psi_\tau(U_{t,\tau}) w_t^c + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} D_t(\widetilde{\alpha}_\tau - \alpha_\tau)\cdot w_t^c \\ &=:\; A_T + R_T. \end{aligned} \tag{61}$$

Note that Lee (2016) implies that $\widetilde{\alpha}_\tau$ is $\sqrt{T}$-consistent for α_τ. Using the standard decomposition $D_t(\widetilde{\alpha}_\tau - \alpha_\tau) = \mathbb{E}(D_t(\widetilde{\alpha}_\tau - \alpha_\tau)|\mathcal{G}_{t-1}) + \widetilde{D}_t(\widetilde{\alpha}_\tau - \alpha_\tau)$, with $\mathbb{E}(\widetilde{D}_t(\widetilde{\alpha}_\tau - \alpha_\tau)|\mathcal{G}_{t-1}) = 0$, we have

$$R_T \;=\; \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathbb{E}(D_t(\widetilde{\alpha}_\tau - \alpha_\tau)|\mathcal{G}_{t-1}) w_t^c + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \widetilde{D}_t(\widetilde{\alpha}_\tau - \alpha_\tau) w_t^c \;=:\; R_{T,1} + R_{T,2}. \tag{62}$$

Since $\mathbb{E}(D_t(\widetilde{\alpha}_\tau - \alpha_\tau)|\mathcal{G}_{t-1}) = F_{t,\tau}(0|\mathcal{G}_{t-1}) - F_{t,\tau}(\widetilde{\alpha}_\tau - \alpha_\tau|\mathcal{G}_{t-1}) = \tau - F_{t,\tau}(\widetilde{\alpha}_\tau - \alpha_\tau|\mathcal{G}_{t-1})$, we have

$$R_{T,1} \;=\; \Big[-\sqrt{T} f_\tau(0)(\widetilde{\alpha}_\tau - \alpha_\tau) + C\cdot(\widetilde{\alpha}_\tau - \alpha_\tau)^2 \Big] \frac{1}{T}\sum_{t=1}^{T} w_t^c \;=\; 0. \tag{63}$$

by construction. Next, for $R_{T,2}$, noting that $|\widetilde{D}_t(\cdot)| \leq 1$, $|w_t^c| \leq 2$, it follows that

$$\mathbb{E}(R_{T,2}^2|\mathcal{G}_{t-1}) \;\leq\; \frac{C}{T}\sum_{t=1}^{T}\mathbb{E}(D_t(\widetilde{\alpha}_\tau - \alpha_\tau)^2|\mathcal{G}_{t-1}) \;=\; o_p(1), \tag{64}$$

and hence $R_T = o_p(1)$. Similarly, on noting boundedness of ψ_τ^2, it follows that $T^{-1}\sum_{t=1}^{T}\widetilde{\xi}_{t,\tau}(\beta_\tau)^2 - T^{-1}\sum_{t=1}^{T}((\psi_\tau(U_{t,\tau})w_t^c)^2 = o_p(1)$.

Since X_{t-1} is $\mathcal{G}_{t-1}$-measurable, we have $\mathbb{E}(\psi_\tau(U_{t,\tau})|\mathcal{G}_{t-1}) = 0$ and $\mathbb{E}(\psi_\tau(U_{t,\tau})^2 \mid \mathcal{G}_{t-1}) = \tau(1-\tau)$. Therefore, $\psi_\tau(U_{t,\tau})$ is a martingale difference with respect to $\mathcal{G}_{t-1}$, and the rest of the proof is closely similar to the proof of Theorem 3. Write

$$\Delta_{Tt} \;:=\; \begin{pmatrix}\psi_\tau(U_{t,\tau})w_t\\ \psi_\tau(U_{t,\tau})\end{pmatrix}, \qquad M_T \;:=\; \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta_{Tt} = \begin{pmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\tau(U_{t,\tau})w_t\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\tau(U_{t,\tau})\end{pmatrix}. \tag{65}$$

Then, the quadratic variation of M_T is given by

$$\langle M\rangle_T := \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\Delta_{Tt}\Delta_{Tt}'|\mathcal{G}_{t-1}) = \tau(1-\tau)\cdot\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}w_t^2 & w_t\\ w_t & 1\end{pmatrix}, \tag{66}$$

and following the same steps in the proof of Theorem 3, it follows that

$$\begin{pmatrix}1 & -\bar{w}_T\end{pmatrix} M_T \;\xrightarrow{st}\; MN(0,\eta^2), \tag{67}$$

where

$$\eta_T^2 := \begin{pmatrix}1 & -\bar{w}_T\end{pmatrix}\langle M\rangle_T\begin{pmatrix}1\\ -\bar{w}_T\end{pmatrix} = \tau(1-\tau)\cdot\frac{1}{T}\sum_{t=1}^{T}(w_t^c)^2 \;\xrightarrow{p}\; \eta^2. \tag{68}$$

In the meantime, since w_t is $\mathcal{G}_{t-1}$-measurable and $\mathbb{E}(\psi_\tau(U_{t,\tau})^2|\mathcal{G}_{t-1}) = \tau(1-\tau)$, we can straightforwardly show consistency of the self-normalizer as before. That is,

$$\begin{aligned} V_T =\; & \frac{1}{T}\sum_{t=1}^{T}\widetilde{\xi}_{t,\tau}(\beta_{\tau,0})^2 \\ =\; & \frac{1}{T}\sum_{t=1}^{T}\psi_\tau(U_{t,\tau})^2 w_t^2 - 2\bar{w}_T\frac{1}{T}\sum_{t=1}^{T}\psi_\tau(U_{t,\tau})^2 w_t + \bar{w}_T^2\frac{1}{T}\sum_{t=1}^{T}\psi_\tau(U_{t,\tau})^2 + o_p(1) \\ =\; & \tau(1-\tau)\cdot\frac{1}{T}\sum_{t=1}^{T}(w_t^c)^2 + o_p(1) \;\xrightarrow{p}\; \eta^2. \end{aligned} \tag{69}$$

Consequently, as $T\to\infty$,

$$\frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{\xi}_{t,\tau}(\beta_{\tau,0})}{\sqrt{V_T}} \;\xrightarrow{st}\; N(0,1), \qquad \text{and hence} \qquad \frac{\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{\xi}_{t,\tau}(\beta_{\tau,0})\right)^2}{V_T} \;\xrightarrow{st}\; \chi_1^2. \tag{70}$$

Therefore, $\widetilde{\ell}_{T,\tau}(\beta_\tau)$ converges in distribution to χ_1^2 as desired. The proof is now complete. □

Supplementary Internet Appendix to accompany "Unified Inference for Predictive Mean and Quantile Regressions via Empirical Likelihood"

Zongwu Cai* Yifeng Chen† Seok Young Hong‡ Daniel Tsvetanov§

January 27, 2026

*University of Kansas, Lawrence, KS 66045, United States of America ; caiz@ku.edu.
†Nanyang Technological University, Singapore 639798, Singapore ; yifeng002@e.ntu.edu.sg.
‡Nanyang Technological University, Singapore 639798, Singapore ; seokyoung.hong@ntu.edu.sg.
§Norwich Business School, University of East Anglia, Norwich NR4 7TJ, UK; d.tsvetanov@uea.ac.uk.

IA.1 Simulation Design under Conditional Heteroscedasticity

In this section, we describe the Monte Carlo simulations set up that we use to investigate the finite sample behavior of the proposed EL methods under the assumption of conditional heteroscedasticity in the innovations of the predictive regression system (Assumption 2 in the paper). The overall structure of the simulation (including persistence levels, endogeneity parameters, sample size, number of replications, and nominal significance level) remains unchanged from the design described in Section 5.1 of the main paper. Specifically, we continue to consider $c \in \{-50, -20, 0, 2\}$ for predictor persistence, $\phi \in \{-0.5, -0.2, 0\}$ for endogeneity, $T = 250$, and 10,000 Monte Carlo replications. The key modification is that the shocks driving the predictor and the predictive regression error follow GARCH(1,1) processes. For each simulation, conditional volatilities $\sigma_{V,t}$ and $\sigma_{z,t}$ evolve according to

$$\sigma_{V,t}^2 = (1 - a_V - b_V) + a_V V_{t-1}^2 + b_V \sigma_{V,t-1}^2, \tag{IA1}$$

$$\sigma_{z,t}^2 = (1 - a_z - b_z) + a_z z_{t-1}^2 + b_z \sigma_{z,t-1}^2, \tag{IA2}$$

with $a_i, b_i \geq 0$ and $a_i + b_i < 1$, where $i \in \{V, z\}$. Inovations are generated as $V_t = \sigma_{V,t}\varepsilon_{V,t}$ and $z_t = \sigma_{z,t}\varepsilon_{z,t}$, where $(\varepsilon_{V,t}, \varepsilon_{z,t})$ is drawn from a bivariate standard normal distribution. Note that $\mathbb{E}\left(V_t^2\right) = \mathbb{E}\left(z_t^2\right) = 1$. Results are reported for $a_V = a_z = 0.05$ and $b_V = b_z = 0.9$. For the mean regression, the dependent variable is constructed as:

$$Y_t = \alpha + \beta X_{t-1} + U_t, \tag{IA3}$$

$$U_t = \phi V_t + z_t. \tag{IA4}$$

with X_t evolving as in the main paper and $\varepsilon_t = V_t$. For quantile regressions, we adjust the error term to enforce the null at quantile τ by subtracting its conditional quantile:

$$U_{t,\tau} = U_t - Q_{U_t}\left(\tau \mid \mathcal{G}_{t-1}\right), \tag{IA5}$$

$$Q_{U_t}\left(\tau \mid \mathcal{G}_{t-1}\right) = \sigma_{U,t}\Phi^{-1}(\tau), \tag{IA6}$$

where $\sigma_{U,t}^2 = \phi^2\sigma_{V,t}^2 + \sigma_{z,t}^2$ and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function (CDF).

All other aspects of the design (including the range of local alternatives, test implementation, and evaluation metrics) follow the homoscedastic case described in the main paper. This extension allows us to assess the robustness of the proposed EL procedures under conditional heteroscedasticity, a feature commonly observed in financial return data. As reported in IA.3, simulation results under heteroscedasticity are qualitatively similar.

IA.2 Additional diagnostic simulations for EL3 under strong endogeneity

This appendix provides additional discussion and diagnostic simulation evidence to understand the mild size distortion of the two-stage empirical likelihood procedure (EL3) in mean predictability tests under strong endogeneity (e.g., $\phi = -0.5$), as documented in Remark 9.

Background: Cai and Wang (2014) consider the model

$$y_t = \beta_0 + \beta_1 u_t + \beta_2 x_{t-1} + v_t, \tag{IA7}$$

$$x_t = \theta + \rho x_{t-1} + u_t, \tag{IA8}$$

and propose a two-step procedure where Step 1 estimates (θ, ρ) by OLS and forms $\widehat{u}_t = x_t - (\widehat{\theta} + \widehat{\rho} x_{t-1})$, and Step 2 estimates $(\beta_0, \beta_1, \beta_2)$ using the regression $y_t = \beta_0 + \beta_1 \widehat{u}_t + \beta_2 x_{t-1} + v_t$. In their setting, v_t is orthogonal to (x_{t-1}, u_t), so the regression error does not contain an innovation-driven component correlated with the regressor. In our predictive mean regression:

$$Y_t = \alpha + \beta X_{t-1} + U_t, \tag{IA9}$$

$$X_t = \theta + \rho X_{t-1} + V_t, \tag{IA10}$$

$$U_t = \phi V_t + z_t, \tag{IA11}$$

the two-stage implementation (EL3) first estimates (θ, ρ) by OLS in the autoregression for X_t, i.e. (IA10), and obtains the residual $\widehat{V}_t$. It then estimates α by OLS from

$$Y_t = \alpha + \gamma \widehat{V}_t + \beta X_{t-1} + \xi_t, \tag{IA12}$$

where γ serves as a proxy for ϕ when $\widehat{V}_t$ is close to V_t. The resulting intercept estimate $\widetilde{\alpha}$ is used to form the EL score in the main paper.

Method 1 (baseline EL3 used in the paper). Our baseline EL3 uses the residual $Y_t - \widetilde{\alpha}$ in the EL constraint, i.e.,

$$\widetilde{Z}_t(\beta) = \left[(Y_t - \widetilde{\alpha}) - \beta X_{t-1}\right] \cdot w^c(X_{t-1}). \tag{IA13}$$

Under the null $\beta = 0$, the residual $Y_t - \widetilde{\alpha}$ is approximately $U_t = \phi V_t + z_t$ and therefore still contains the innovation-driven component ϕV_t when $\phi \neq 0$. When X_t is highly persistent, the component $\phi V_t \, w^c(X_{t-1})$ may induce stronger dependence and heavier tails in the moment condition, leading to mild over-rejection in finite samples (see Table IA5). Intuitively, strong

persistence makes $w(X_{t-1})$ remain close to its saturation levels for long stretches, so the centering term $\bar{w}_T$ can be dominated by a few long runs, and $w^c(X_{t-1}) = w(X_{t-1}) - \bar{w}_T$ becomes highly uneven over time (often near zero but occasionally large). Multiplying this uneven centered weight by the innovation component ϕV_t can generate clustered variability and occasional large realizations in the score, which inflates higher-order moments and weakens finite-sample calibration.

Method 2 (A feasible purging using estimated innovations). A natural modification is to "purge" the innovation-driven component using the estimated innovation, and use the residual $Y_t - \widetilde{\alpha} - \widetilde{\phi}\widehat{V}_t$ so that the EL score becomes

$$\widetilde{Z}_t(\beta) = \left[(Y_t - \widetilde{\alpha} - \widetilde{\phi}\widehat{V}_t) - \beta X_{t-1}\right] \cdot w^c(X_{t-1}). \tag{IA14}$$

where $\widetilde{\phi}$ is obtained from (IA12). This approach reduces the innovation component in the score by construction and can ameliorate over-rejection in highly persistent cases. However, since $\widehat{V}_t = V_t + (\widehat{\rho} - \rho)X_{t-1} + (\widehat{\theta} - \theta)$, the purged residual contains the additional term $\widetilde{\phi}(\widehat{\rho} - \rho)X_{t-1}$. Under stationarity (or near stationarity), $\widehat{\rho} - \rho$ converges more slowly than highly persistent cases, so this estimation-error component is more pronounced in finite samples and can introduce size distortions in stationary or local-to-unity regimes. By contrast, under unit-root or mildly explosive persistence, the convergence rates for $\widehat{\rho}$ is generally faster, so the same contamination term tends to be smaller, explaining why the feasible purging performs better in highly persistent designs.

Method 3 (oracle purging using the true innovation). As an oracle benchmark, we consider purging based on the true innovation V_t and use the residual

$$Y_t - \widetilde{\alpha} - \widetilde{\phi}V_t \tag{IA15}$$

in the EL score. This oracle approach eliminates the size distortion across persistence classes in our simulations, as evidenced in Table IA5, confirming that the mild over-rejection of the baseline EL3 under strong endogeneity is primarily driven by the innovation-correlated component that remains in the residual when V_t is not purged.

Because V_t is unobserved in practice, feasible purging based on $\widehat{V}_t$ involves a trade-off: it can reduce the innovation-driven component that matters under strong endogeneity and high persistence, but it may introduce additional estimation-error effects in stationary or near-stationary regimes. We therefore report the baseline EL3 (Method 1) results in the paper and document these diagnostics in Table IA5.

IA.3 Additional Simulation Results

Table IA1
Finite-Sample Sizes for Mean Predictive Regressions with Homoscedastic Errors and no Intercept in the DGP for X_t

The table reports the empirical size (i.e., the probability of incorrectly rejecting the null hypothesis of no predictability). Results are shown for: EL1 (the EL method, where the intercept α is treated as known; Section 2.1), EL2 (sample-splitting EL approach in Section 2.2), EL3 (the two-stage EL procedure using the projection method in Section 2.3, and IVX (benchmark test of Kostakis et al. (2015) and Phillips and Lee (2016)). All tests are conducted at the 5% nominal significance level. EL tests are conducted using the hyperbolic tangent weights $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The simulation design accounts for various levels of persistence in the predictor X_t through the localizing constant $c \in \{-50, -20, 0, 2\}$, and endogeneity through the innovation correlation parameter $\phi \in \{-0.50, -0.20, 0\}$. The data is generated as described in Section 5.1 of the main paper, with $\theta = 0$ in the DGP for X_t. Rejection probabilities are based on 10,000 Monte Carlo simulations and sample size $T = 250$.

c	$\phi = 0$				$\phi = -0.2$				$\phi = -0.5$			
	EL1	EL2	EL3	IVX	EL1	EL2	EL3	IVX	EL1	EL2	EL3	IVX
-50	0.0533	0.0528	0.0546	0.0534	0.0535	0.0526	0.0544	0.0546	0.0545	0.0515	0.0560	0.0513
-20	0.0523	0.0497	0.0521	0.0545	0.0540	0.0508	0.0546	0.0531	0.0547	0.0498	0.0582	0.0543
0	0.0462	0.0503	0.0475	0.0477	0.0492	0.0516	0.0544	0.0494	0.0524	0.0581	0.0827	0.0486
2	0.0494	0.0560	0.0523	0.0526	0.0498	0.0556	0.0574	0.0544	0.0547	0.0613	0.0904	0.0553

Table IA2
Finite-Sample Sizes for Quantile Regressions with Homoscedastic Errors and no Intercept in the DGP for X_t

The table reports the empirical size (i.e., the probability of incorrectly rejecting the null hypothesis of no predictability) in the quantile regression model (13) of the main paper, for a selected set of quantiles $\tau \in \{0.1, 0.2, \ldots, 0.9\}$. Results are shown for EL1 (the EL method, where the intercept α_τ is treated as known; Section 3.1), EL3 (the two-stage EL procedure), and IVX (benchmark test of Lee (2016)). All tests are conducted at the 5% nominal significance level. EL tests are conducted using the hyperbolic tangent weights $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The simulation design accounts for various levels of persistence in the predictor X_t through the localizing constant $c \in \{-50, -20, 0, 2\}$, and endogeneity through the innovation correlation parameter $\phi \in \{-0.50, -0.20, 0\}$. The data is generated as described in Section 5.1 of the main paper, with $\theta = 0$ in the DGP for X_t. Rejection probabilities are based on 10,000 Monte Carlo simulations and sample size $T = 250$.

c	τ	$\phi = 0$			$\phi = -0.2$			$\phi = -0.5$		
		EL1	EL3	IVX	EL1	EL3	IVX	EL1	EL3	IVX
-50	0.1	0.0578	0.0699	0.0650	0.0582	0.0706	0.0688	0.0556	0.0646	0.0629
	0.2	0.0544	0.0588	0.0513	0.0542	0.0607	0.0525	0.0545	0.0596	0.0432
	0.3	0.0523	0.0556	0.0450	0.0525	0.0588	0.0473	0.0503	0.0555	0.0432
	0.4	0.0504	0.0559	0.0404	0.0496	0.0539	0.0405	0.0490	0.0534	0.0379
	0.5	0.0490	0.0519	0.0381	0.0480	0.0537	0.0363	0.0503	0.0556	0.0374
	0.6	0.0486	0.0558	0.0400	0.0490	0.0543	0.0398	0.0468	0.0545	0.0360
	0.7	0.0498	0.0537	0.0457	0.0493	0.0543	0.0473	0.0478	0.0571	0.0361
	0.8	0.0496	0.0549	0.0520	0.0478	0.0549	0.0475	0.0527	0.0570	0.0487
	0.9	0.0588	0.0672	0.0693	0.0586	0.0671	0.0667	0.0531	0.0639	0.0612
-20	0.1	0.0563	0.0680	0.0654	0.0553	0.0663	0.0679	0.0559	0.0669	0.0624
	0.2	0.0525	0.0572	0.0536	0.0556	0.0589	0.0509	0.0556	0.0633	0.0443
	0.3	0.0515	0.0551	0.0470	0.0532	0.0582	0.0461	0.0507	0.0593	0.0402
	0.4	0.0506	0.0538	0.0440	0.0482	0.0518	0.0438	0.0502	0.0573	0.0386
	0.5	0.0481	0.0532	0.0431	0.0493	0.0516	0.0393	0.0492	0.0567	0.0374
	0.6	0.0510	0.0538	0.0366	0.0497	0.0549	0.0408	0.0487	0.0566	0.0361
	0.7	0.0474	0.0535	0.0423	0.0467	0.0547	0.0454	0.0481	0.0557	0.0386
	0.8	0.0463	0.0548	0.0501	0.0512	0.0569	0.0483	0.0501	0.0589	0.0446
	0.9	0.0580	0.0690	0.0659	0.0541	0.0637	0.0641	0.0539	0.0638	0.0602
0	0.1	0.0519	0.0548	0.1570	0.0536	0.0579	0.1530	0.0528	0.0666	0.0988
	0.2	0.0471	0.0516	0.1344	0.0496	0.0517	0.1295	0.0534	0.0644	0.0732
	0.3	0.0477	0.0508	0.1303	0.0459	0.0538	0.1178	0.0500	0.0670	0.0693
	0.4	0.0463	0.0458	0.1280	0.0457	0.0512	0.1150	0.0471	0.0650	0.0649
	0.5	0.0482	0.0464	0.1243	0.0492	0.0512	0.1105	0.0474	0.0666	0.0638
	0.6	0.0468	0.0488	0.1229	0.0458	0.0502	0.1118	0.0522	0.0667	0.0622
	0.7	0.0474	0.0457	0.1254	0.0497	0.0493	0.1154	0.0503	0.0648	0.0644
	0.8	0.0488	0.0475	0.1353	0.0506	0.0532	0.1282	0.0526	0.0647	0.0753
	0.9	0.0473	0.0550	0.1529	0.0486	0.0569	0.1497	0.0516	0.0660	0.1034
2	0.1	0.0519	0.0584	0.2092	0.0539	0.0609	0.2062	0.0526	0.0726	0.1717
	0.2	0.0504	0.0514	0.1975	0.0491	0.0564	0.1947	0.0557	0.0650	0.1310
	0.3	0.0473	0.0518	0.1827	0.0483	0.0496	0.1749	0.0531	0.0665	0.1216
	0.4	0.0479	0.0478	0.1739	0.0490	0.0485	0.1703	0.0529	0.0664	0.1140
	0.5	0.0486	0.0450	0.1701	0.0502	0.0498	0.1598	0.0523	0.0692	0.1160
	0.6	0.0459	0.0466	0.1687	0.0474	0.0483	0.1603	0.0505	0.0652	0.1129
	0.7	0.0477	0.0494	0.1723	0.0500	0.0500	0.1739	0.0535	0.0720	0.1230
	0.8	0.0479	0.0510	0.1881	0.0501	0.0548	0.1923	0.0517	0.0679	0.1334
	0.9	0.0493	0.0582	0.2092	0.0490	0.0600	0.2121	0.0533	0.0720	0.1647

Table IA3
Finite-Sample Sizes for Mean Predictive Regressions with Heteroscedastic Errors

The table reports the empirical size (i.e., the probability of incorrectly rejecting the null hypothesis of no predictability). Results are shown for: EL1 (the EL method, where the intercept α is treated as known; Section 2.1), EL2 (sample-splitting EL approach in Section 2.2), EL3 (the two-stage EL procedure using the projection method in Section 2.3, and IVX (benchmark test of Kostakis et al. (2015) and Phillips and Lee (2016)). All tests are conducted at the 5% nominal significance level. EL tests are conducted using the hyperbolic tangent weights $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The simulation design accounts for various levels of persistence in the predictor X_t through the localizing constant $c \in \{-50, -20, 0, 2\}$, and endogeneity through the innovation correlation parameter $\phi \in \{-0.50, -0.20, 0\}$. The data is generated as described in Section IA.1, allowing for conditional heteroscedasticity in the innovations of the predictive regression system. Rejection probabilities are based on 10,000 Monte Carlo simulations and sample size $T = 250$.

c	$\phi = 0$				$\phi = -0.2$				$\phi = -0.5$			
	EL1	EL2	EL3	IVX	EL1	EL2	EL3	IVX	EL1	EL2	EL3	IVX
-50	0.0526	0.0535	0.0527	0.0554	0.0537	0.0528	0.0560	0.0552	0.0545	0.0539	0.0557	0.0536
-20	0.0536	0.0496	0.0533	0.0539	0.0537	0.0507	0.0552	0.0562	0.0549	0.0504	0.0588	0.0572
0	0.0464	0.0481	0.0457	0.0477	0.0479	0.0520	0.0541	0.0489	0.0539	0.0564	0.0860	0.0507
2	0.0490	0.0540	0.0531	0.0532	0.0505	0.0537	0.0584	0.0530	0.0558	0.0603	0.0914	0.0559

Table IA4
Finite-Sample Sizes for Quantile Regressions with Heteroscedastic Errors

The table reports the empirical size (i.e., the probability of incorrectly rejecting the null hypothesis of no predictability) in the quantile regression model (12), for a selected set of quantiles $\tau \in \{0.1, 0.2, \ldots, 0.9\}$. Results are shown for EL1 (the EL method, where the intercept α_τ is treated as known), EL3 (the two-stage EL procedure), and IVX (benchmark test of Lee (2016)). All tests are conducted at the 5% nominal significance level. EL tests are conducted using the hyperbolic tangent weights $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The simulation design accounts for various levels of persistence in the predictor X_t through the localizing constant $c \in \{-50, -20, 0, 2\}$, and endogeneity through the innovation correlation parameter $\phi \in \{-0.50, -0.20, 0\}$. The data is generated as described in Section IA.1 allowing for conditional heteroscedasticity in the innovations of the predictive regression system. Rejection probabilities are based on 10,000 Monte Carlo simulations and sample size $T = 250$.

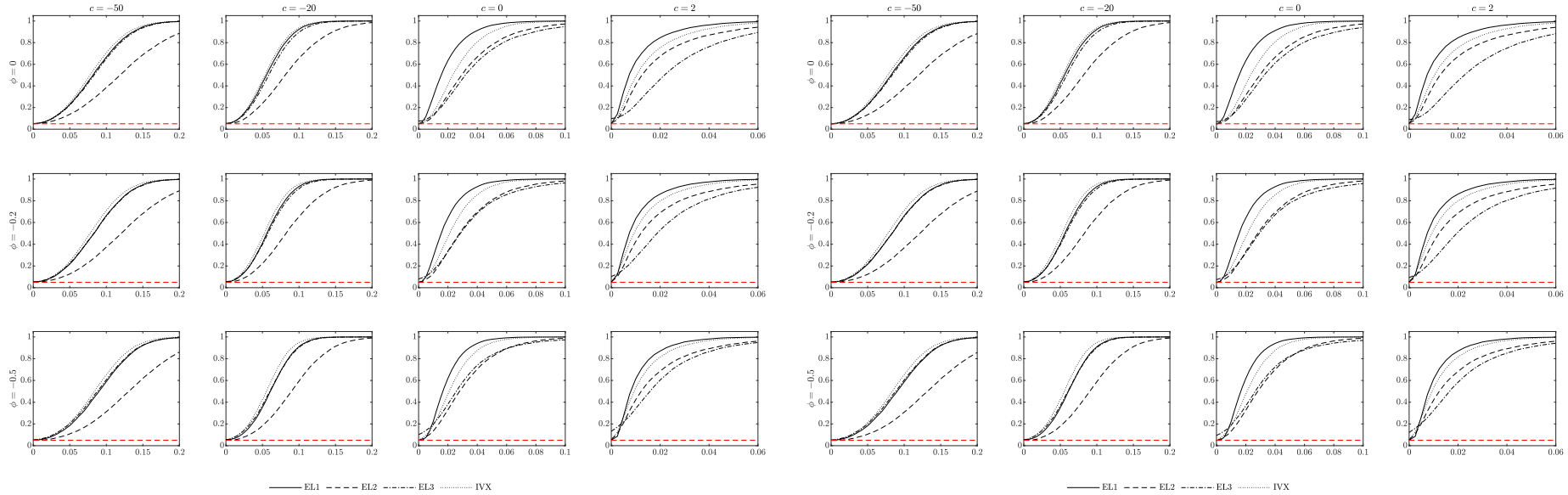
c	τ	$\phi = 0$			$\phi = -0.2$			$\phi = -0.5$		
		EL1	EL3	IVX	EL1	EL3	IVX	EL1	EL3	IVX
-50	0.1	0.0587	0.0674	0.0651	0.0602	0.0715	0.0670	0.0578	0.0659	0.0668
	0.2	0.0545	0.0589	0.0549	0.0544	0.0607	0.0539	0.0528	0.0609	0.0476
	0.3	0.0524	0.0567	0.0446	0.0505	0.0575	0.0484	0.0510	0.0557	0.0467
	0.4	0.0502	0.0555	0.0409	0.0506	0.0546	0.0426	0.0486	0.0541	0.0406
	0.5	0.0494	0.0523	0.0402	0.0481	0.0512	0.0408	0.0519	0.0563	0.0370
	0.6	0.0499	0.0547	0.0432	0.0505	0.0551	0.0407	0.0486	0.0539	0.0385
	0.7	0.0490	0.0541	0.0470	0.0487	0.0561	0.0474	0.0500	0.0550	0.0425
	0.8	0.0491	0.0570	0.0510	0.0499	0.0567	0.0519	0.0528	0.0553	0.0505
	0.9	0.0598	0.0686	0.0714	0.0588	0.0683	0.0651	0.0554	0.0658	0.0692
-20	0.1	0.0558	0.0693	0.0693	0.0560	0.0651	0.0685	0.0558	0.0653	0.0665
	0.2	0.0540	0.0574	0.0539	0.0533	0.0587	0.0517	0.0543	0.0626	0.0480
	0.3	0.0510	0.0555	0.0487	0.0541	0.0583	0.0474	0.0486	0.0592	0.0423
	0.4	0.0509	0.0533	0.0468	0.0473	0.0500	0.0418	0.0492	0.0566	0.0414
	0.5	0.0480	0.0523	0.0431	0.0481	0.0521	0.0432	0.0495	0.0571	0.0387
	0.6	0.0519	0.0532	0.0366	0.0498	0.0549	0.0410	0.0498	0.0552	0.0382
	0.7	0.0481	0.0543	0.0484	0.0460	0.0524	0.0464	0.0487	0.0573	0.0397
	0.8	0.0469	0.0544	0.0511	0.0496	0.0580	0.0495	0.0524	0.0583	0.0497
	0.9	0.0569	0.0691	0.0662	0.0558	0.0672	0.0674	0.0535	0.0641	0.0670
0	0.1	0.0530	0.0562	0.1582	0.0527	0.0573	0.1554	0.0558	0.0666	0.1054
	0.2	0.0480	0.0511	0.1380	0.0494	0.0530	0.1303	0.0536	0.0643	0.0782
	0.3	0.0474	0.0517	0.1331	0.0476	0.0547	0.1191	0.0519	0.0665	0.0728
	0.4	0.0475	0.0492	0.1334	0.0482	0.0530	0.1164	0.0486	0.0670	0.0705
	0.5	0.0480	0.0500	0.1256	0.0465	0.0548	0.1137	0.0486	0.0694	0.0668
	0.6	0.0463	0.0514	0.1255	0.0455	0.0518	0.1147	0.0498	0.0678	0.0689
	0.7	0.0469	0.0498	0.1272	0.0488	0.0507	0.1180	0.0513	0.0640	0.0717
	0.8	0.0481	0.0485	0.1345	0.0495	0.0550	0.1276	0.0508	0.0669	0.0752
	0.9	0.0472	0.0555	0.1544	0.0472	0.0562	0.1483	0.0520	0.0673	0.1062
2	0.1	0.0532	0.0599	0.2094	0.0541	0.0602	0.2092	0.0552	0.0720	0.1718
	0.2	0.0503	0.0523	0.1959	0.0504	0.0563	0.1911	0.0554	0.0690	0.1363
	0.3	0.0479	0.0529	0.1841	0.0489	0.0520	0.1758	0.0548	0.0668	0.1256
	0.4	0.0504	0.0488	0.1775	0.0507	0.0526	0.1728	0.0523	0.0676	0.1217
	0.5	0.0478	0.0471	0.1701	0.0498	0.0523	0.1654	0.0551	0.0716	0.1219
	0.6	0.0472	0.0487	0.1705	0.0479	0.0538	0.1629	0.0505	0.0701	0.1243
	0.7	0.0478	0.0522	0.1724	0.0497	0.0540	0.1756	0.0527	0.0739	0.1268
	0.8	0.0497	0.0530	0.1883	0.0493	0.0574	0.1918	0.0534	0.0707	0.1384
	0.9	0.0494	0.0616	0.2098	0.0491	0.0606	0.2110	0.0544	0.0701	0.1677

Table IA5
EL3 size across Method 1–3 in Section IA.2

The table reports the empirical size of the test statistics computed using Method 1 (baseline EL3), Method 2 (the feasible purging), Method 3 (the oracle purging) described in Section IA.2 of the Internet Appendix, for selected combinations of (ϕ, c) with $T = 250$, $\theta = 0 = \phi$, and tanh weighting. Results for Method 1 correspond to relevant entries of Table IA1.

ϕ	c	Method 1	Method 2	Method 3
-0.50	-50	0.0560	0.0817	0.0560
-0.50	-20	0.0582	0.0822	0.0531
-0.50	0	0.0827	0.0653	0.0476
-0.50	2	0.0904	0.0675	0.0527
0.00	-50	0.0546	0.0545	0.0562
0.00	-20	0.0521	0.0527	0.0531
0.00	0	0.0475	0.0475	0.0491
0.00	2	0.0523	0.0526	0.0534

Figure IA1
Finite-Sample Size and Power Plots for Mean Predictability Tests, with Homoscedastic Errors and Conventional Weights

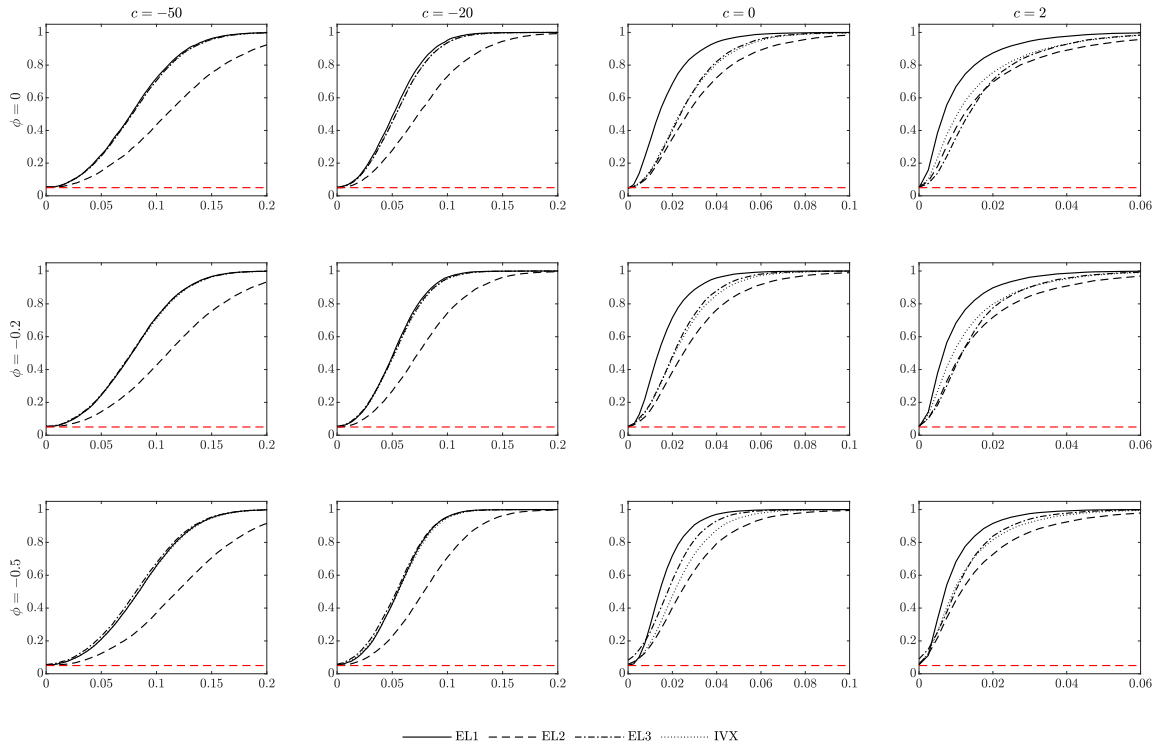


Panel A: Without Bartlett Correction

Panel B: With Bartlett Correction

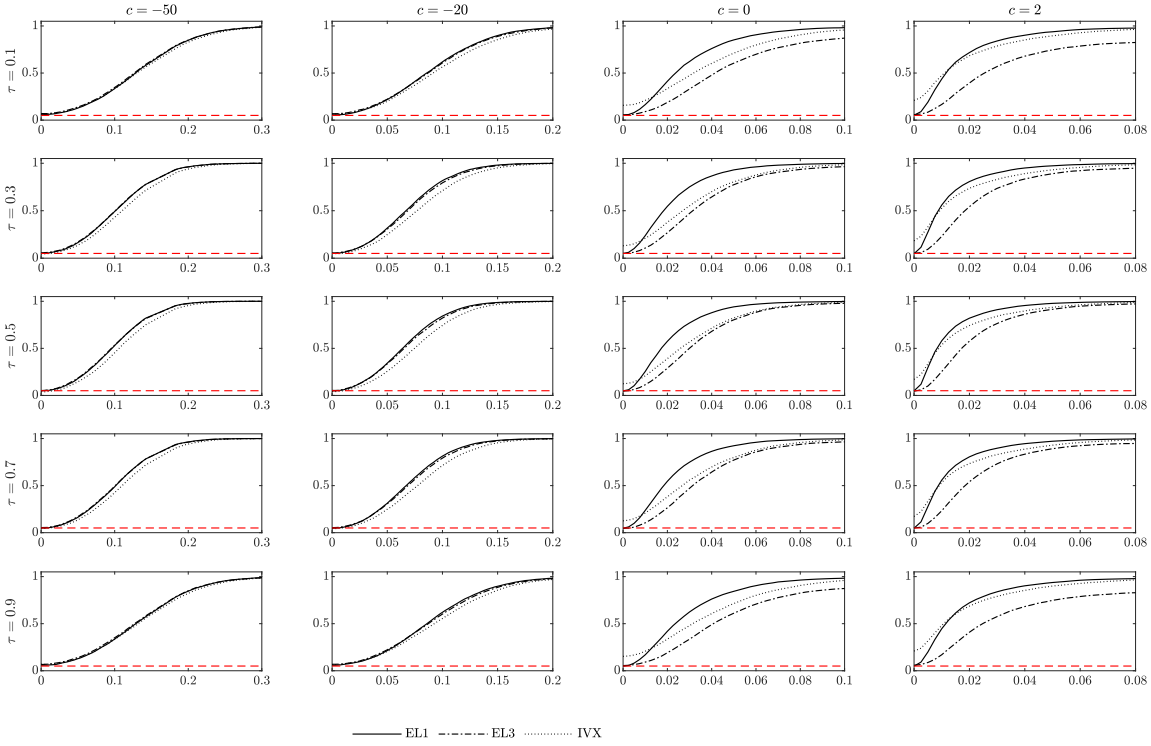
The figure summarizes rejection probabilities (y-axis) for tests of the null hypothesis of no predictive ability in the mean regression model (1) of the main paper, with $\theta = 0$. Results are shown for: EL1 (the EL method, where the intercept α is treated as known; Section 2.1), EL2 (sample-splitting EL approach in Section 2.2), EL3 (the two-stage EL procedure using the projection method in Section 2.3, and IVX (benchmark test of Kostakis et al. (2015) and Phillips and Lee (2016)). All EL tests are conducted using the conventional weights function $w(X_{t-1}) = X_{t-1}/\sqrt{1 + X_{t-1}^2}$, without and with Bartlett correction (Panel A and B, respectively). The x-axis represents true values of the slope coefficient β, with $\beta = 0$ corresponding to empirical size. The red dashed line marks the 5% nominal level. Rejection probabilities are based on 10,000 Monte Carlo simulations (see Section 5.1 of the main paper for details of the simulation design). Results are reported for sample size $T = 250$, innovation correlation parameter $\phi \in \{-0.5, -0.2, 0\}$, and localizing constant $c \in \{-50, -20, 0, 2\}$.

Figure IA2
Finite-Sample Size and Power Plots for Mean Predictability Tests, with Homoscedastic Errors and Hyperbolic Tangent Weight



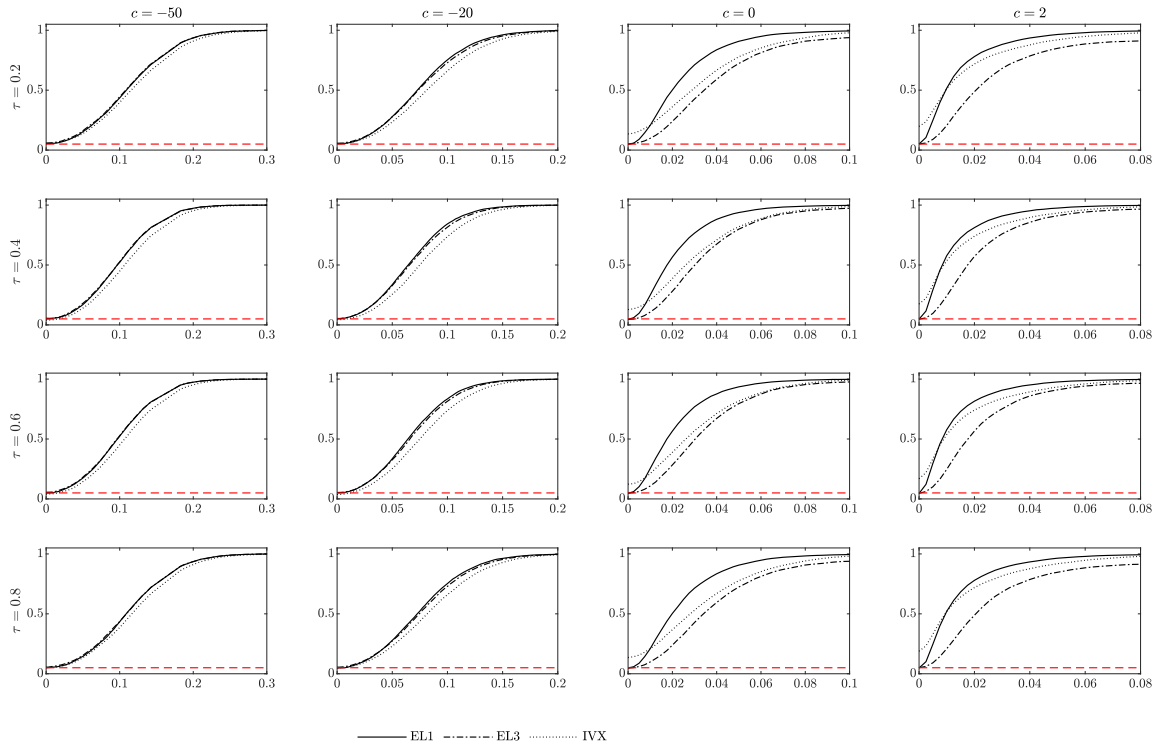
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the mean regression model (1), and letting $\theta = 0$ in the DGP for X_t. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The figure under the same setting with $\theta = 0.05$ is Figure 2 in the main paper. See also notes to Figure IA1.

Figure IA3
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Homoscedastic Errors and Correlation $\phi = 0$



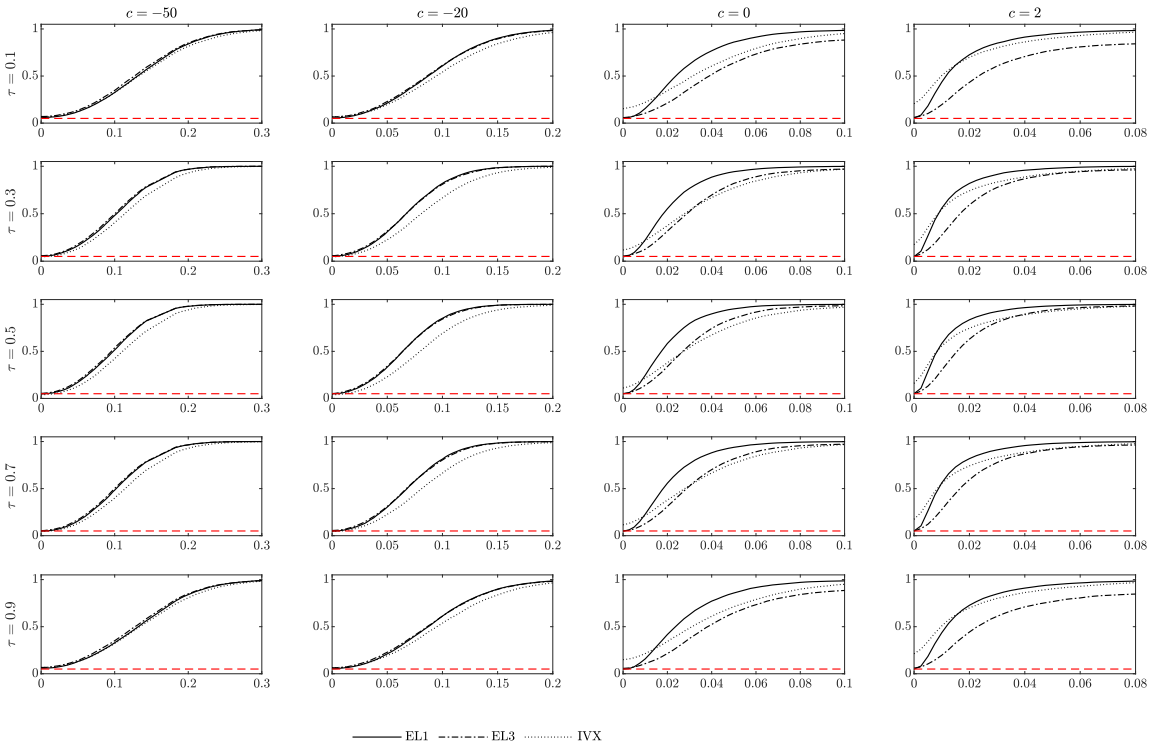
The figure summarizes rejection probabilities (y-axis) for tests of the null hypothesis of no predictive ability in the quantile regression model (13) in the main paper, for a selected set of quantiles $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. Results are shown for EL1 (the EL method, where the intercept α_τ is treated as known), EL3 (the two-stage EL procedure outlined in Section 3.2), and IVX (benchmark test of Lee (2016)). EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying Bartlett correction. The x-axis represents true values of the slope coefficient β_τ, with $\beta_\tau = 0$ corresponding to empirical size. The red dashed line marks the 5% nominal level. Rejection probabilities are based on 10,000 Monte Carlo simulations (see Section 5.1 of the main paper for full details of the simulation design). Results are reported for sample size $T = 250$, localizing constant $c \in \{-50, -20, 0, 2\}$, and innovation correlation parameter $\phi = 0$.

Figure IA4
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Homoscedastic Errors and Correlation $\phi = 0$



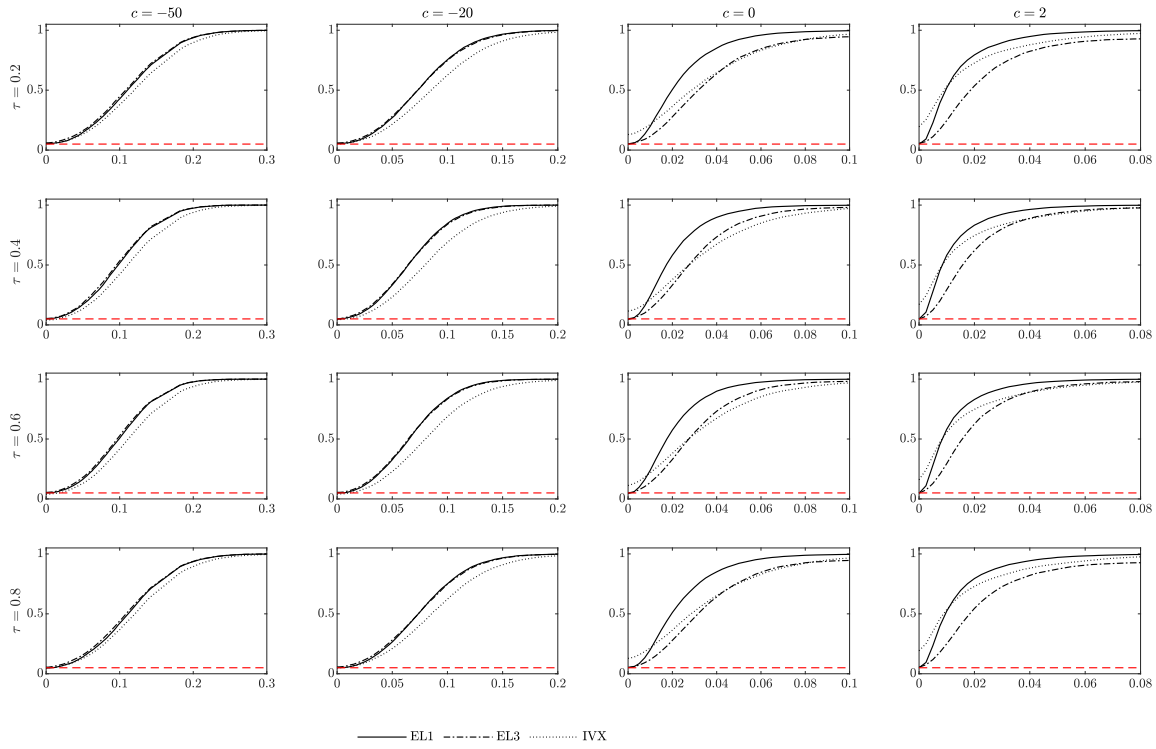
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13) of the main paper, for a selected set of quantiles $\tau \in \{0.2, 0.4, 0.6, 0.8\}$, and innovation correlation parameter $\phi = 0$. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. See also notes to Figure IA3.

Figure IA5
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Homoscedastic Errors and Correlation $\phi = -0.2$



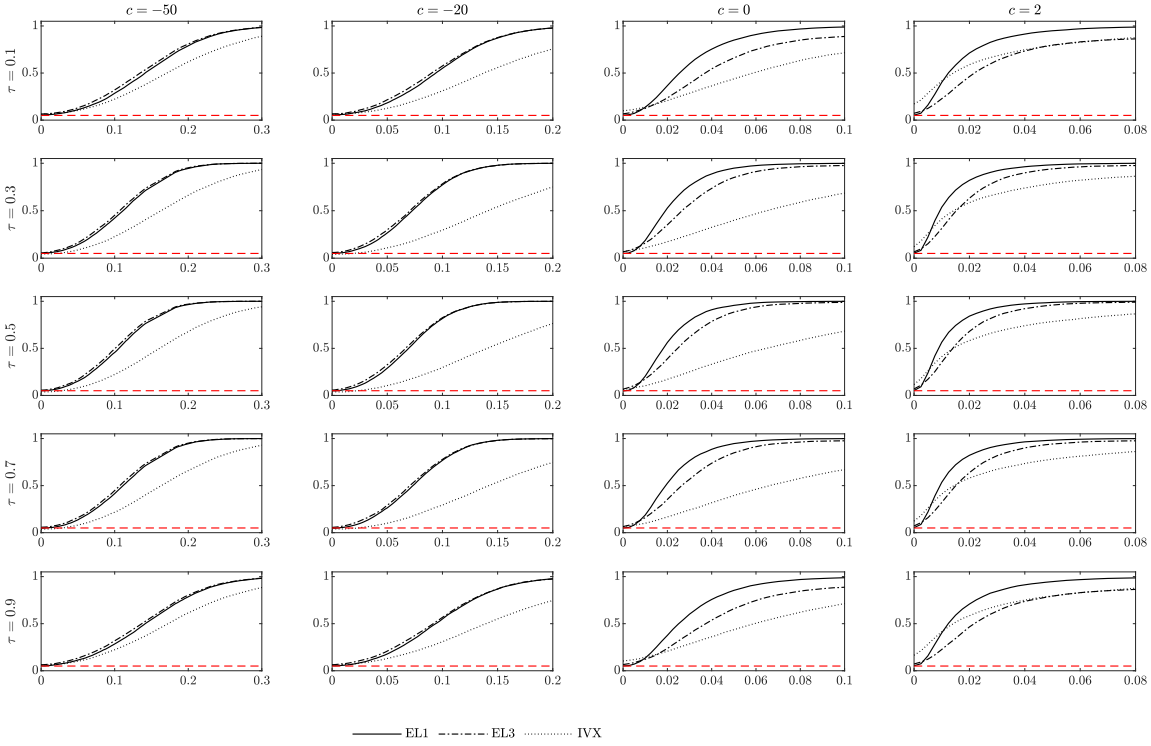
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13) of the main paper, for a selected set of quantiles $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, and innovation correlation parameter $\phi = -0.2$. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. See also notes to Figure IA3.

Figure IA6
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Homoscedastic Errors and Correlation $\phi = -0.2$



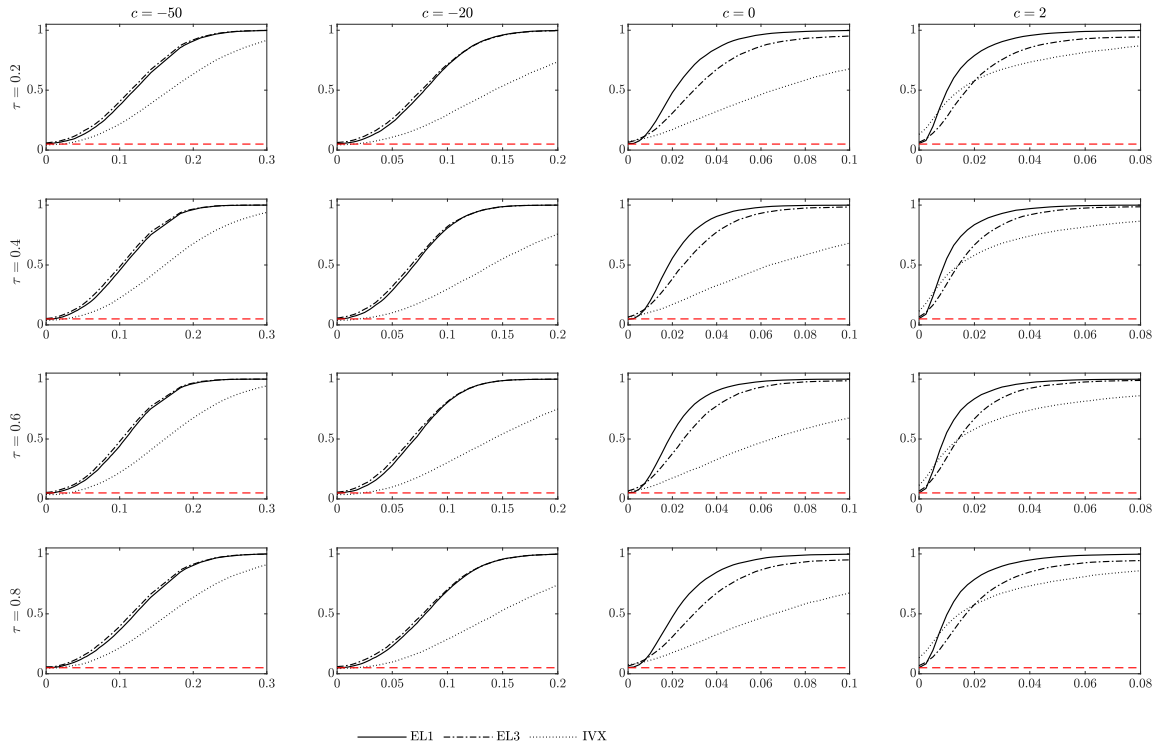
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13) of the main paper, for a selected set of quantiles $\tau \in \{0.2, 0.4, 0.6, 0.8\}$, and innovation correlation parameter $\phi = -0.2$. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. See also notes to Figure IA3.

Figure IA7
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Homoscedastic Errors and Correlation $\phi = -0.5$



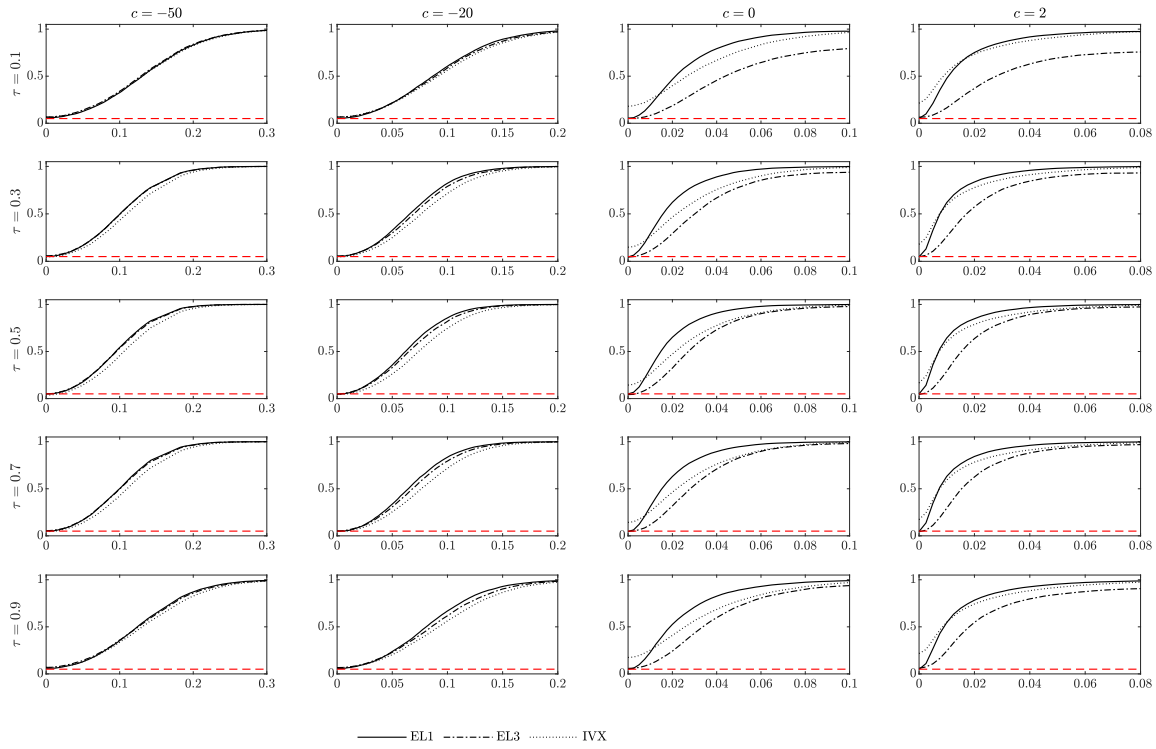
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13) of the main paper, for a selected set of quantiles $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, and innovation correlation parameter $\phi = -0.5$. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. See also notes to Figure IA3.

Figure IA8
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Homoscedastic Errors and Correlation $\phi = -0.5$



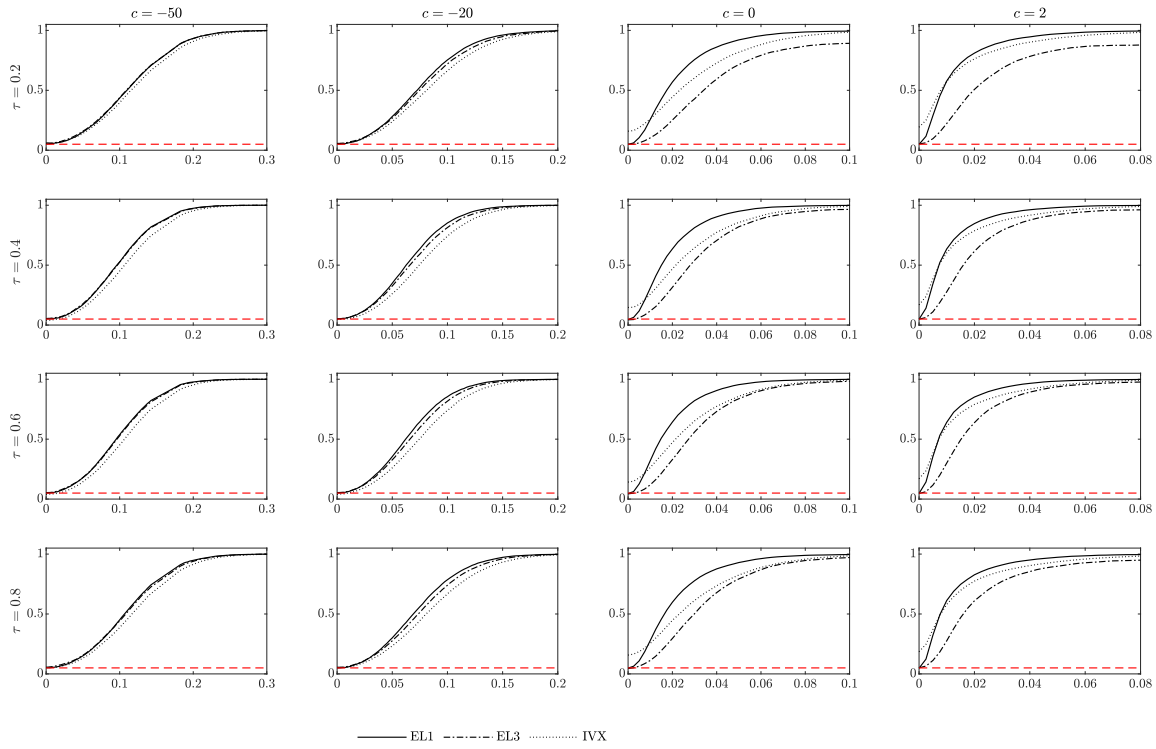
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13) of the main paper, for a selected set of quantiles $\tau \in \{0.2, 0.4, 0.6, 0.8\}$, and innovation correlation parameter $\phi = -0.5$. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. See also notes to Figure IA3.

Figure IA9
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Homoscedastic Errors, Nonzero Intercept in the DGP for X_t, and Correlation $\phi = 0$



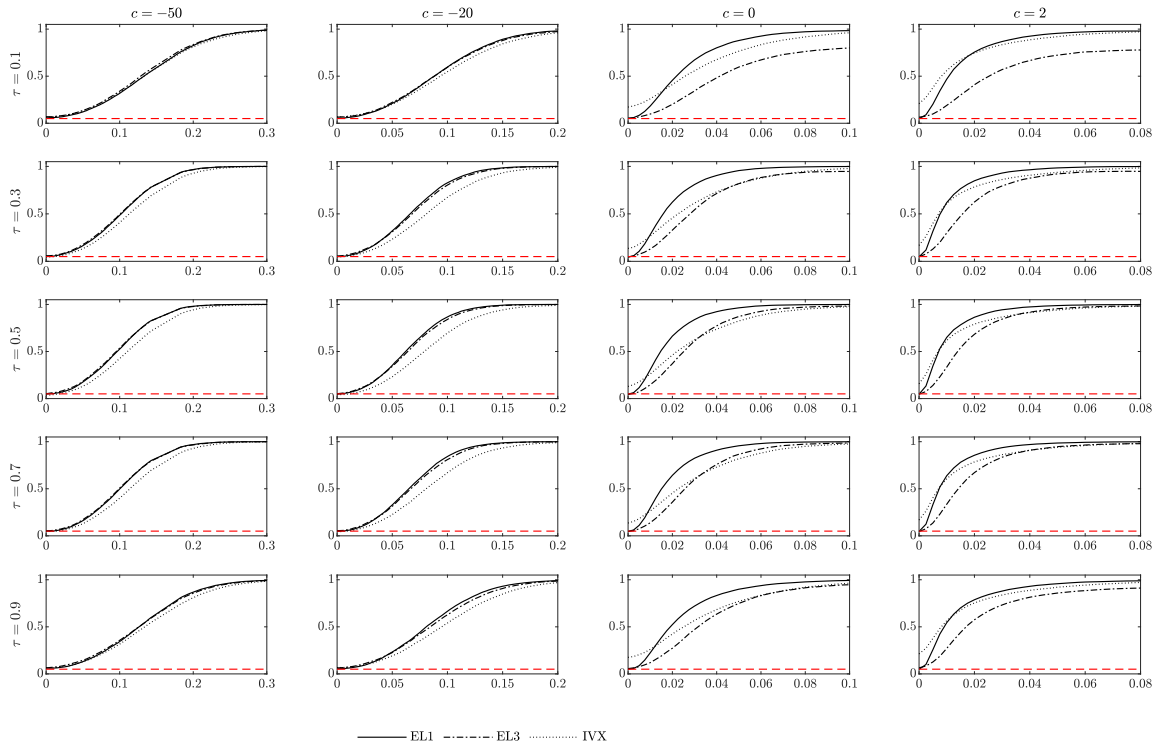
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13) of the main paper, for a selected set of quantiles $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, innovation correlation parameter $\phi = 0$, and $\theta = 0.05$ in the DGP for X_t. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. See also notes to Figure IA3.

Figure IA10
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Homoscedastic Errors, Nonzero Intercept in the DGP for X_t, and Correlation $\phi = 0$



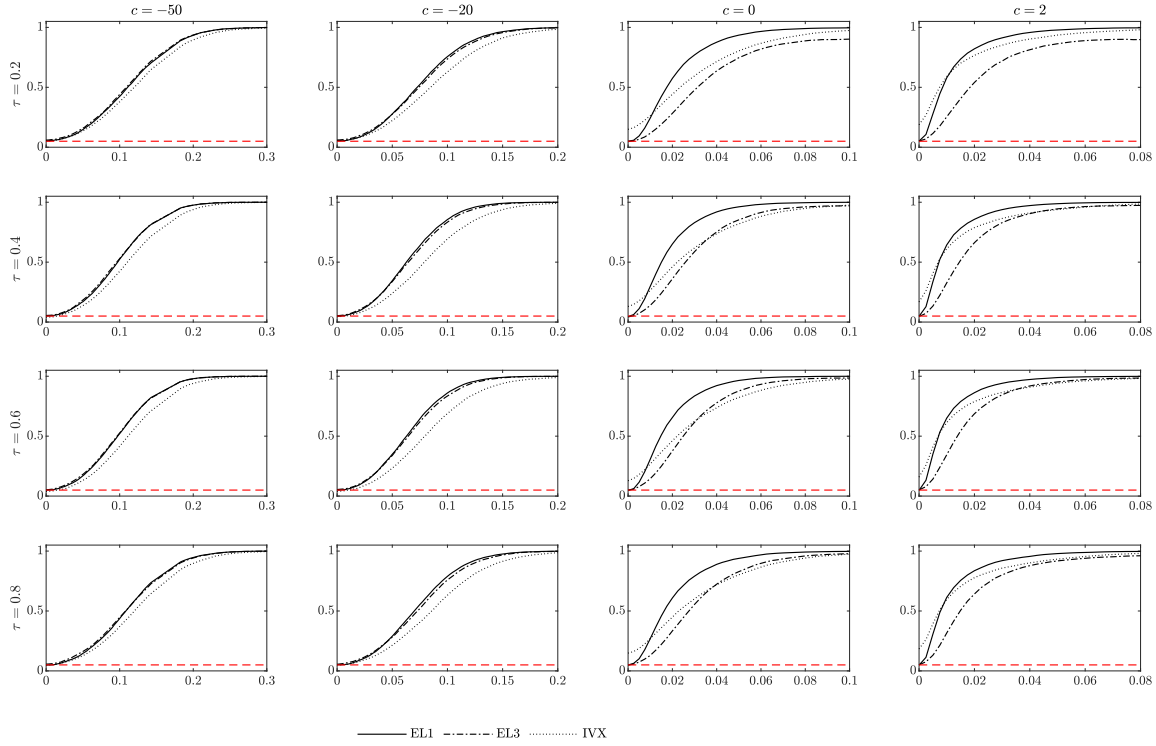
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13) of the main paper, for a selected set of quantiles $\tau \in \{0.2, 0.4, 0.6, 0.8\}$, innovation correlation parameter $\phi = 0$, and $\theta = 0.05$ in the DGP for X_t. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. See also notes to Figure IA3.

Figure IA11
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Homoscedastic Errors, Nonzero Intercept in the DGP for X_t, and Correlation $\phi = -0.2$



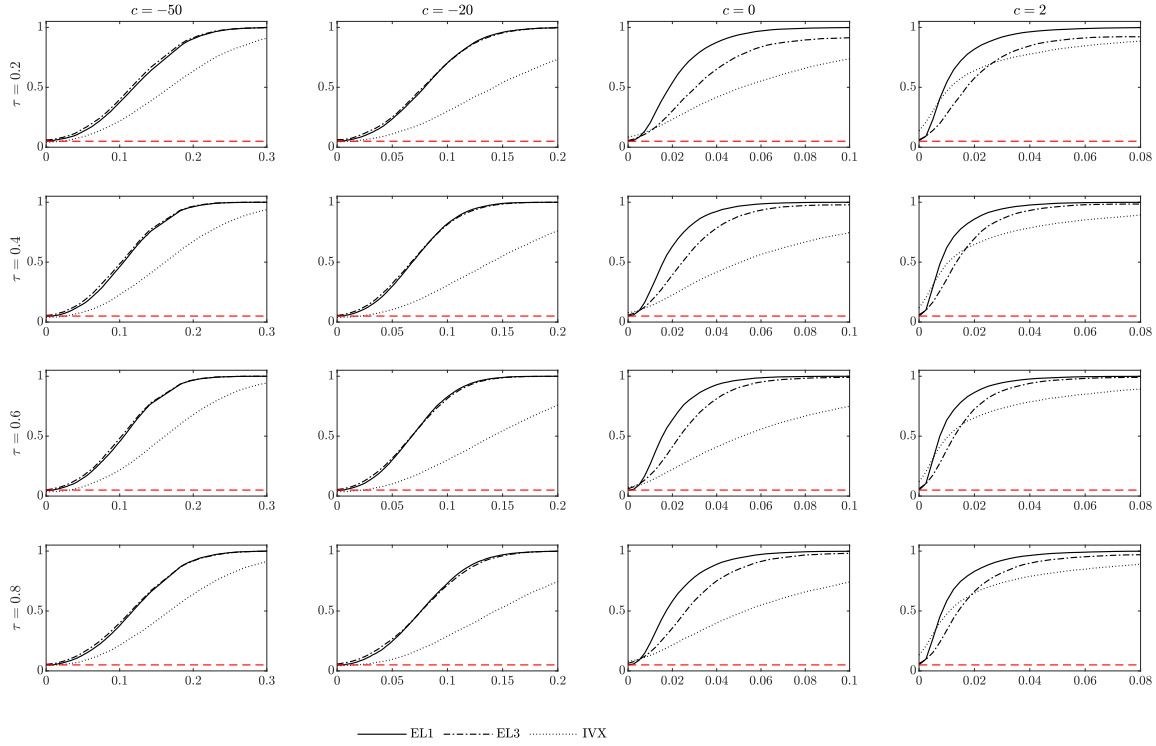
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13) of the main paper, for a selected set of quantiles $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, innovation correlation parameter $\phi = -0.2$, and $\theta = 0.05$ in the DGP for X_t. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. See also notes to Figure IA3.

Figure IA12
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Homoscedastic Errors, Nonzero Intercept in the DGP for X_t, and Correlation $\phi = -0.2$



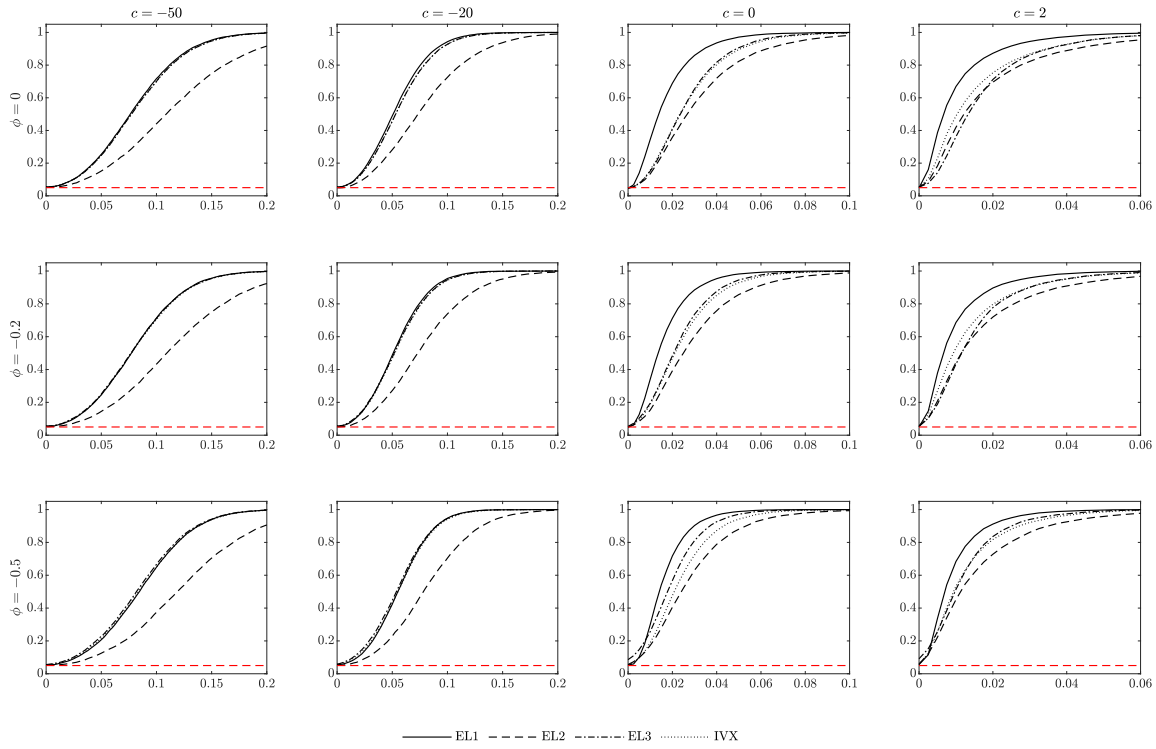
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13) of the main paper, for a selected set of quantiles $\tau \in \{0.2, 0.4, 0.6, 0.8\}$, innovation correlation parameter $\phi = -0.2$, and $\theta = 0.05$ in the DGP for X_t. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. See also notes to Figure IA3.

Figure IA13
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Homoscedastic Errors, Nonzero Intercept in the DGP for X_t, and Correlation $\phi = -0.5$



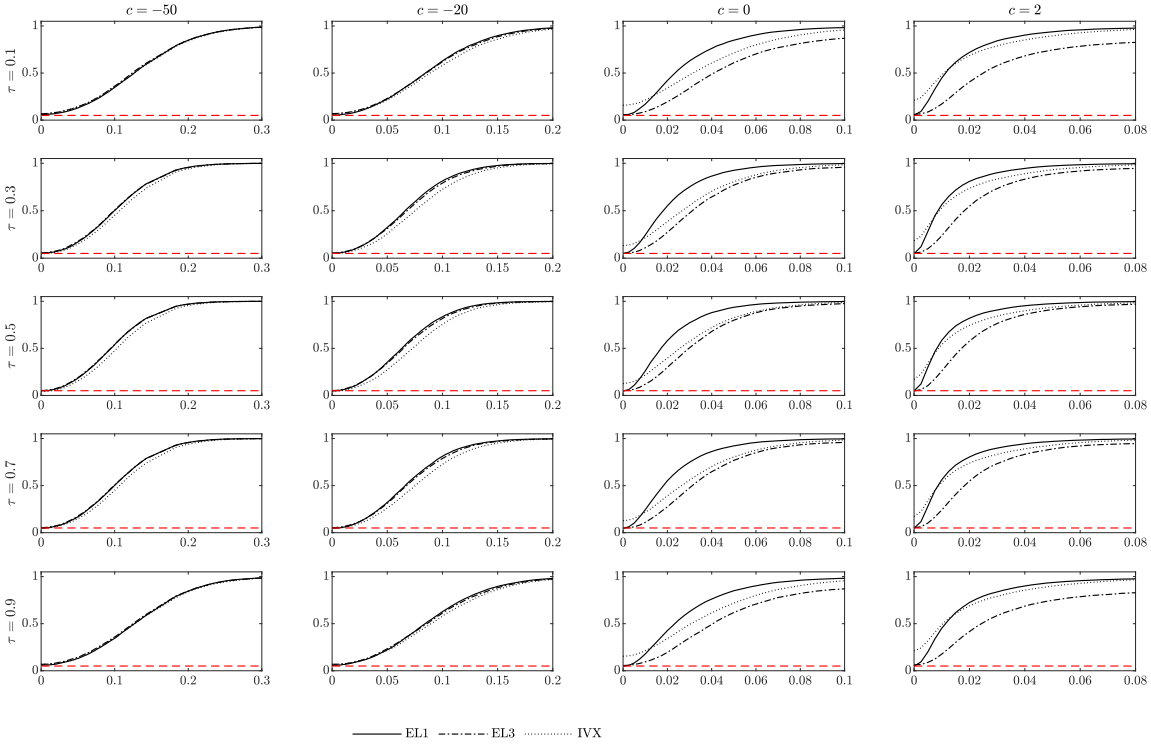
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13) of the main paper, for a selected set of quantiles $\tau \in \{0.2, 0.4, 0.6, 0.8\}$, innovation correlation parameter $\phi = -0.5$, and $\theta = 0.05$ in the DGP for X_t. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. See also notes to Figure IA3. The figure under the same setting with the set of quantiles $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ is Figure 3 in the main paper.

Figure IA14
Finite-Sample Size and Power Plots for Mean Predictability Tests, with Heteroscedastic Errors



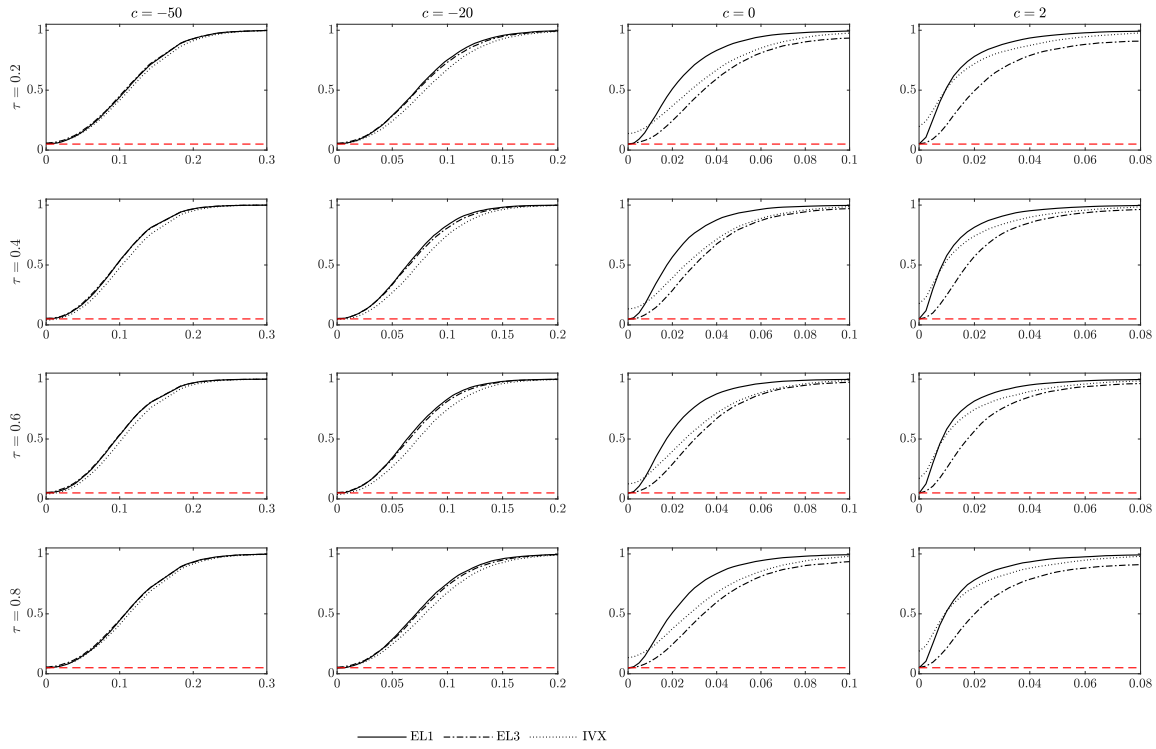
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the mean regression model (1), and letting $\theta = 0$ in the DGP for X_t. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The simulation design is as described in Section IA1, allowing for heteroscedasticity in the innovations of the predictive regression system. Results are reported for sample size $T = 250$, innovation correlation parameter $\phi \in \{-0.5, -0.2, 0\}$, and localizing constant $c \in \{-50, -20, 0, 2\}$. See also notes to Figure IA1. The figure under the same setting with homoscedastic error and $\theta = 0$ is Figure 2 in the main paper.

Figure IA15
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Heteroscedastic Errors and Correlation $\phi = 0$



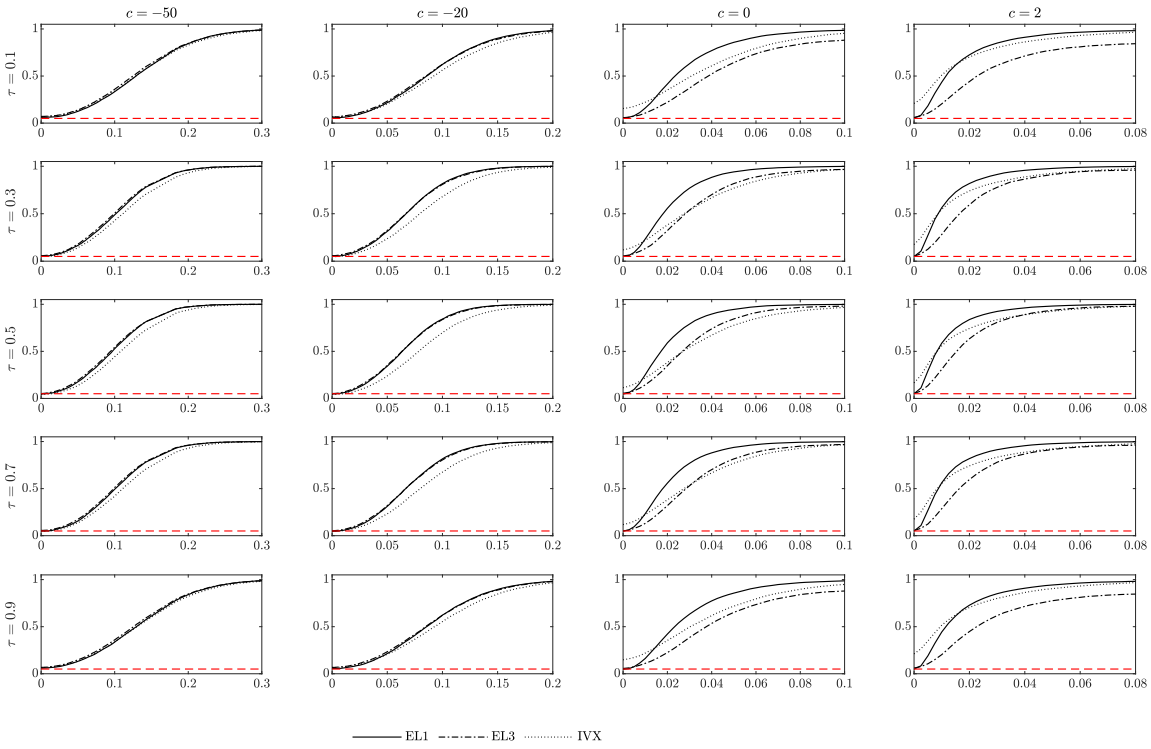
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13), for a selected set of quantiles $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The simulation design is as described in Section IA1, allowing for heteroscedasticity in the innovations of the predictive regression system. Results are reported for sample size $T = 250$, localizing constant $c \in \{-50, -20, 0, 2\}$, and innovation correlation parameter $\phi = 0$. See also notes to Figure IA3.

Figure IA16
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Heteroscedastic Errors and Correlation $\phi = 0$



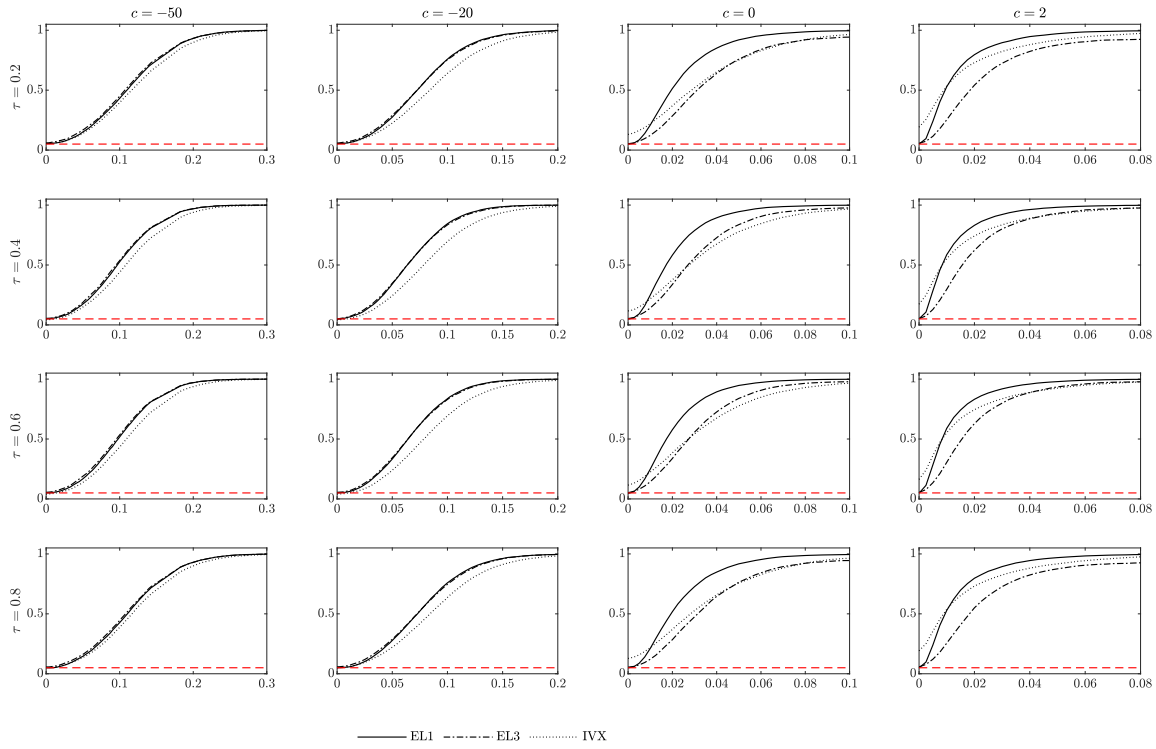
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13), for a selected set of quantiles $\tau \in \{0.2, 0.4, 0.6, 0.8\}$. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The simulation design is as described in Section IA1, allowing for heteroscedasticity in the innovations of the predictive regression system. Results are reported for sample size $T = 250$, localizing constant $c \in \{-50, -20, 0, 2\}$, and innovation correlation parameter $\phi = 0$. See also notes to Figure IA3.

Figure IA17
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Heteroscedastic Errors and Correlation $\phi = -0.2$



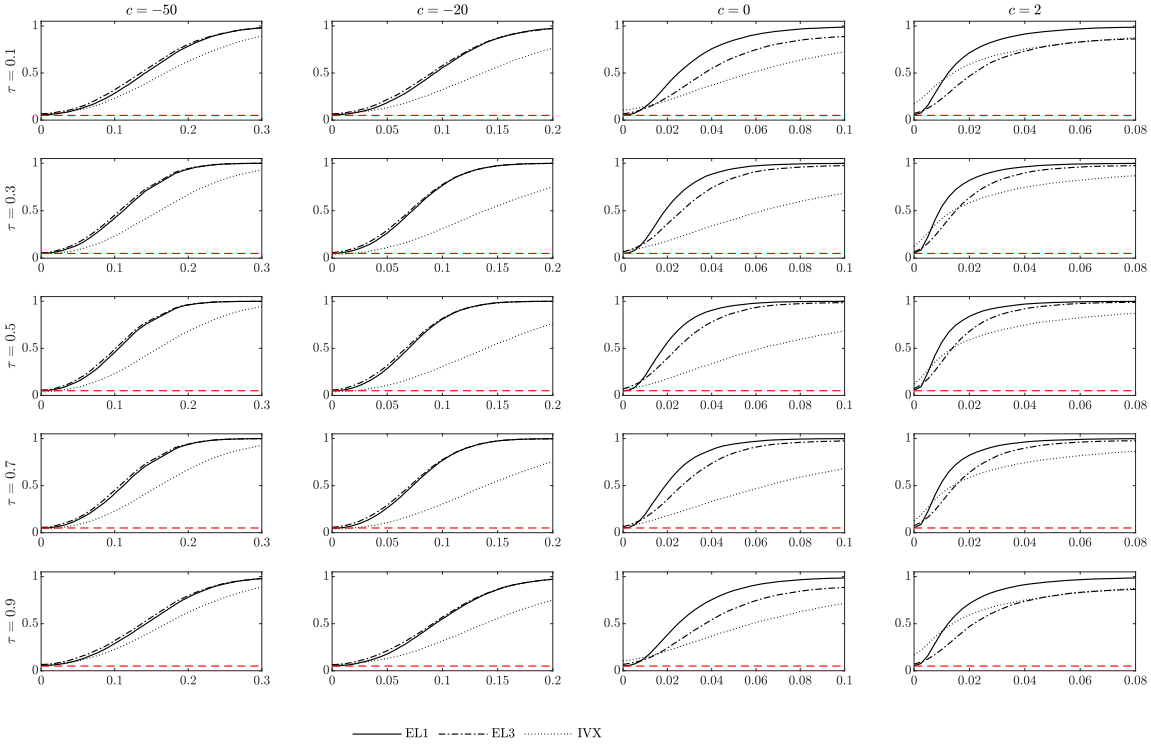
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13), for a selected set of quantiles $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The simulation design is as described in Section IA1, allowing for heteroscedasticity in the innovations of the predictive regression system. Results are reported for sample size $T = 250$, localizing constant $c \in \{-50, -20, 0, 2\}$, and innovation correlation parameter $\phi = -0.2$. See also notes to Figure IA3.

Figure IA18
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Heteroscedastic Errors and Correlation $\phi = -0.2$



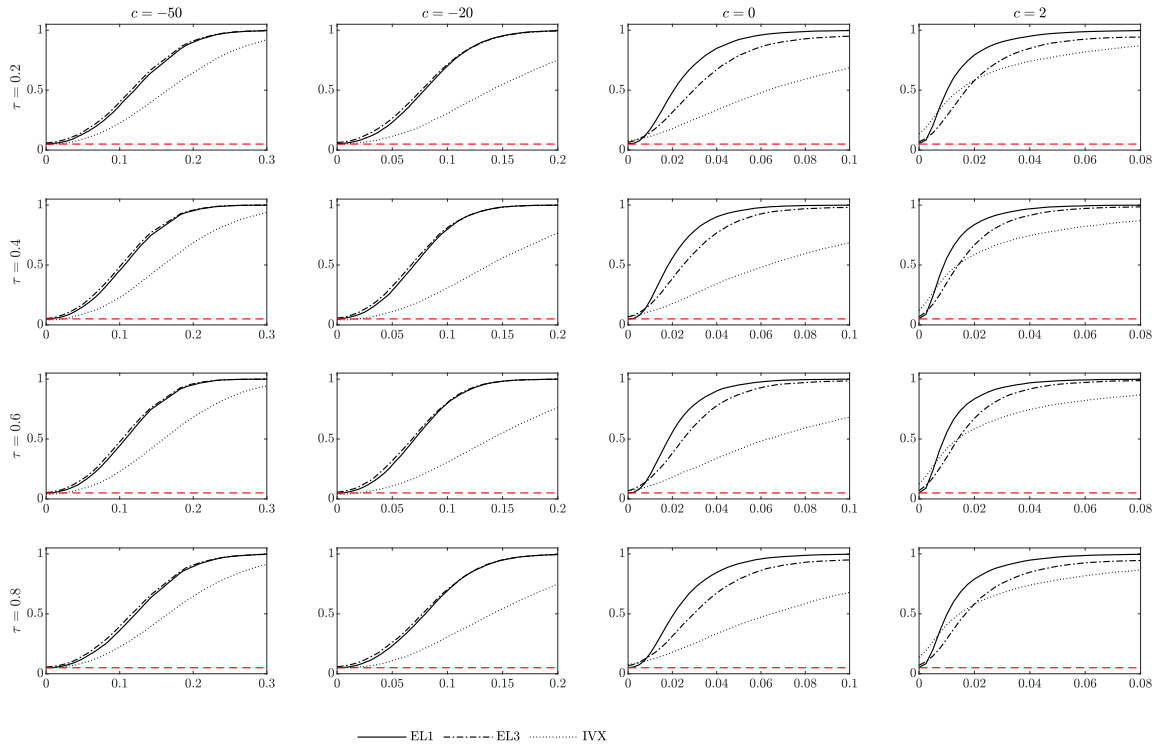
The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13), for a selected set of quantiles $\tau \in \{0.2, 0.4, 0.6, 0.8\}$. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The simulation design is as described in Section IA1, allowing for heteroscedasticity in the innovations of the predictive regression system. Results are reported for sample size $T = 250$, localizing constant $c \in \{-50, -20, 0, 2\}$, and innovation correlation parameter $\phi = -0.2$. See also notes to Figure IA3.

Figure IA19
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Heteroscedastic Errors and Correlation $\phi = -0.5$



The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13), for a selected set of quantiles $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The simulation design is as described in Section IA1, allowing for heteroscedasticity in the innovations of the predictive regression system. Results are reported for sample size $T = 250$, localizing constant $c \in \{-50, -20, 0, 2\}$, and innovation correlation parameter $\phi = -0.5$. See also notes to Figure IA3.

Figure IA20
Finite-Sample Size and Power Plots for Quantile Predictability Tests, with Heteroscedastic Errors and Correlation $\phi = -0.5$



The figure summarizes rejection probabilities for tests of the null hypothesis of no predictive ability in the quantile regression model (13), for a selected set of quantiles $\tau \in \{0.2, 0.4, 0.6, 0.8\}$. All EL tests are conducted using the hyperbolic tangent weights function $w(X_{t-1}) = \tanh(X_{t-1})/10$ and applying the Bartlett correction. The simulation design is as described in Section IA1, allowing for heteroscedasticity in the innovations of the predictive regression system. Results are reported for sample size $T = 250$, localizing constant $c \in \{-50, -20, 0, 2\}$, and innovation correlation parameter $\phi = -0.5$. See also notes to Figure IA3.

References

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

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

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

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