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

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

September 29, 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, Taisuke Otsu, Guangming Pan, Liang Peng, Patrick Pun, Tatevik Sekhposyan, Myunghwan Matt Seo, Won-Ki Seo, Dakyung Seong, Zhentao Shi, Sorawoot Srisuma, Wenjie Wang, Yoon-Jae Whang, and seminar participants at the Chinese University of Hong Kong, Kobe University, University of Illinois Urbana-Champaign, University of Kansas, University of Oxford, and the National University of Singapore. We are also grateful to the participants of the 2025 Midwest Econometrics Group (MEG) Conference, the 2025 Econometric Society World Congress in Seoul, the 2026 Econometric Society Conferences in Cairo and Hong Kong, SETA 2026 in Tokyo, 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 other suitable types of response and possibly persistent predictor in a predictive regression. We note that while we set Y_t and X_{t-1} to be scalars below, multivariate extensions are feasible.

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

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

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

where U_t and ε_t are error terms satisfying regularity conditions that are specified shortly. We

impose the initial condition of $X_0 = O_p(1)$, in line with Phillips (2015) and standard practice in the literature. The unknown slope coefficient β captures the predictive content of X_{t-1} for the conditional mean of Y_t. In this paper, β is interpreted as a predictive coefficient and is not assigned a causal structural interpretation. The main object of interest is to test:

$$H_0 : \beta = \beta_0 \tag{3}$$

where β_0 is a fixed hypothesized value. In many applications, the null hypothesis of interest is no mean predictability, which is obtained by setting $\beta_0 = 0$ so that $H_0 : \beta = 0$.

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:*

$$U_t = \phi\varepsilon_t + z_t, \tag{4}$$

where (z_t, ε_t) is a stationary, ergodic martingale difference with respect to the filtration $\mathcal{G}_t = \sigma(X_0, \{z_s, \varepsilon_s; s \leq t\})$ and satisfies $\sup_t \mathbb{E}(|\varepsilon_t|^{2+q} + |z_t|^{2+q}) < \infty$ for some $q > 0$. Moreover, ε_t is either conditionally homoscedastic or conditionally heteroscedastic, with stationary and strictly positive conditional variance. The process (z_t, ε_t) satisfy $\mathbb{E}(z_t\varepsilon_t) = 0$.

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, ε_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, 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 remark that ergodicity will need to be strengthened to α-mixing when sample-splitting is employed for case C1, see Theorem 2; the result under C2-C5 in Theorem 2 remains valid under ergodicity alone. Also, note that the orthogonality condition $\mathbb{E}(\varepsilon_t z_t) = 0$ is not required for the asymptotic results below. It is imposed simply to conveniently interpret ϕ as the coefficient from the linear projection of U_t on ε_t, i.e. $\phi = \mathbb{E}(U_t\varepsilon_t)/\mathbb{E}(\varepsilon_t^2)$.

Assumption 2. *The conditional variance of the error term U_t is finite and bounded away from zero, a.s. for every t. That is, there exists a constant $c > 0$ such that*

$$\mathbb{E}(U_t^2|\mathcal{G}_{t-1}) = \sigma_t^2 \geq c \tag{5}$$

almost surely for every t. We write $\sigma^2 := \mathbb{E}(U_1^2) \in (0, \infty)$.

Remark 2. Assumption 2 is a mild condition that allows for conditional heteroscedasticity and includes conditional homoscedasticity as a special case, where $\sigma_t^2 = \sigma^2$ almost surely for all t. Similar conditions appear in Assumption A2 of Cai et al. (2023); see also Park (2002) for related discussions. We later show in the appendix (cf. (41)) that Assumptions 1 and 2 imply

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(U_t^2 \mid \mathcal{G}_{t-1}) \xrightarrow{p} \sigma^2 \tag{6}$$

as $T \to \infty$. The assumption does not require stationarity of σ_t^2 and accommodates certain forms its nonstationarity, such as processes with certain structural break structure 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. 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

$$\rho_T \;=\; 1 + \frac{c}{T^a}, \tag{7}$$

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 $w : \mathbb{R} \to \mathbb{R}$ is globally Lipschitz, uniformly bounded, $w(x) \to 1$ as $x \to \infty$, $w(x) \to -1$ as $x \to -\infty$, and there exist constants $C < \infty$ and $\gamma \geq 1$ such that $|w(x) - \mathrm{sgn}(x)| \leq C|x|^{-\gamma}$, $|x| \geq 1$. As we discuss below, we can show that these mild conditions are satisfied by a wide range of weight functions used in the literature. Hereafter, we suppose that the weight function we consider satisfies these conditions.

Some examples of such weight functions include

$$w_1(x) = \frac{x}{\sqrt{1+x^2}}, \quad w_2(x) = \frac{x}{1+|x|}, \quad w_3(x) = \tanh(x/b), \tag{8}$$

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. Note that for C1, we assume throughout that $\mathrm{Var}(w(X_0')) > 0$, where X_t' is the stationary solution of (2); this is a natural condition that holds automatically for the weights in (8) under nondegenerate innovations.

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

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

The profile empirical likelihood function for β is given by:

$$L_T(\beta) \;=\; \sup\left\{ \prod_{t=1}^{T} (T\pi_t) :\; \pi_t \geq 0 \text{ for all } t, \sum_{t=1}^{T} \pi_t = 1, \quad \sum_{t=1}^{T} \pi_t Z_t(\beta) = 0 \right\}, \tag{9}$$

where the supremum is taken over the π's. Accordingly, (9) 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:

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

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

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

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 2 are satisfied. Then, under $H_0 : \beta = \beta_0$,*

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

as $T \to \infty$, where β_0 is a fixed hypothesized value.

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 $t = 1, \ldots, m$, where $m = m_T \to \infty$ is the number of split pairs, hence sample-splitting. The rate for m depends on the time series nature of U_t. Since in Assumptions 1 and 2 we considerably relax the i.i.d assumption on U_t in Zhu et al. (2014), where m can be $O(T)$, we require a stronger condition. Specifically, in Theorem 2 below, we let $m = \lfloor T^\kappa \rfloor$ for some $0 < \kappa < 1$, where under C2 we additionally require $\kappa < a$. 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

$$Y_t^* = \beta X_{t-1}^* + U_t^*, \quad \text{and} \quad X_t^* = \rho X_{t-1}^* + \varepsilon_t^* \tag{10}$$

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 (8) is

used, then $w(X^*_{t-1}) = X^*_{t-1}/\sqrt{1+X^{*,2}_{t-1}}$. 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 Assumptions 1 and 2 hold; only under C1, we strengthen ergodicity in Assumption 1 to α-mixing with rate $e^{-c_0 k}$ for some constant $c_0 > 0$ Then, under $H_0 : \beta = \beta_0$ we have*

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

as $T \to \infty$, where β_0 is a fixed hypothesized value.

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. Furthermore, sample splitting ameliorates endogeneity-amplified persistence via differencing, and thus do not yield the first-order asymptotic distortion in the two-stage method, see below.

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:

$$w^c(x) \;:=\; w(x) \;-\; \frac{1}{T}\sum_{s=1}^{T} w(X_{s-1}), \tag{11}$$

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

$$\widetilde{Z}_t(\beta) \;=\; \left[\widetilde{Y}_t - \beta X_{t-1}\right] \cdot w^c(X_{t-1}), \tag{12}$$

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

$$\widetilde{\ell}_T(\beta) \;=\; -2\log \widetilde{L}_T\left(\beta\right) \;=\; 2\sum_{t=1}^{T} \log\Big(1 + \widetilde{\lambda}\widetilde{Z}_t(\beta)\Big), \tag{13}$$

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

$$\sum_{t=1}^{T} \frac{\widetilde{Z}_t(\beta)}{1 + \widetilde{\lambda}\widetilde{Z}_t(\beta)} = 0. \tag{14}$$

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 2 are satisfied. Then, under $H_0 : \beta = \beta_0$ we have*

$$\widetilde{\ell}_T(\beta_0) \xrightarrow{d} \begin{cases} \chi^2_1 & \text{under C1–C2,} \\ \left(\sqrt{1-\lambda_U^2}\, Z + \lambda_U \mathcal{K}\right)^2 & \text{under C3-C5.} \end{cases}$$

as $T \to \infty$, where β_0 is a fixed hypothesized value, $\lambda_U := \mathbb{E}[U_t\varepsilon_t]/(\sigma\sigma_\epsilon)$, and $Z \sim N(0,1)$ is independent of $\mathcal{K}$, a persistence-specific random functional defined in the appendix.

Corollary 1. *In addition to what is assumed in Theorem 3, if $\mathbb{E}[U_t\varepsilon_t] = 0$, then*

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

as $T \to \infty$, under all persistence classes C1, C2, C3, C4, and C5.

The results above can be used to test the hypothesis $H_0 : \beta = \beta_0$ Furthermore, an empirical likelihood confidence set for β can be obtained.

Remark 5. Theorem 3 and Corollary 1 shows how the existence of embdeed endogeneity, i.e. $\phi > 0$, affects the chi-squared distributional limit of the empirical likelihood ratio statistic. Specifically, in the simulation design $U_t = \phi\varepsilon_t + z_t$ with independent standard normal innovations, we have

$$\mathbb{E}[U_t\varepsilon_t] = \phi, \qquad \lambda_U = \frac{\phi}{\sqrt{1+\phi^2}}.$$

Consequently, by Theorem 3, under C3-C5 the limiting law becomes

$$\widetilde{\ell}_T(\beta_0) \Rightarrow \frac{(Z + \phi\mathcal{K})^2}{1+\phi^2}.$$

The zero-covariance condition is therefore automatically satisfied when $\phi = 0$, whereas nonzero ϕ introduces contemporaneous endogeneity and affects the asymptotic distribution. The limiting law approaches χ^2_1 as $\phi \to 0$, thereby characterizing how the chi-squared benchmark is perturbed as endogeneity is introduced.

The noncentrality component provides a theoretical mechanism consistent with the relatively modest over-rejection observed for cases C3 - C5 in the two-stage procedure in Section 5. The magnitude and direction of the distortion depend on the persistence-specific functional $\mathcal{K}$; the limit does not imply that nonzero endogeneity necessarily produces only small over-rejection in every setting. In particular, for fixed nonzero ϕ, departures from the chi-squared calibration need not disappear as the sample size increases.

Remark 6. The two-stage method uses the full sample for inference, thereby circumventing the efficiency loss in sample splitting. Section 4 discusses further finite-sample issues that may arise in the two-stage method due to higher order expansions and proposes remedies via a Bartlett-type 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.

Let $\{\mathcal{F}_t\}$ be a filtration to which $X_t, U_{t,\tau}$, and ε_t are adapted, with $X_{t-1} \in \mathcal{F}_{t-1}$. For a given quantile level $\tau \in (0,1)$, let $Q_\tau(Y_t \mid \mathcal{F}_{t-1})$ denote the conditional τ-quantile of Y_t given $\mathcal{F}_{t-1}$, namely, $\inf\{x : \mathbb{P}(Y_t \le x \mid \mathcal{F}_{t-1}) \ge \tau\}$. Following Lee (2016), Fan and Lee (2019), and Cai et al. (2023), we consider the linear quantile specification:

$$Q_\tau(Y_t \mid \mathcal{F}_{t-1}) = \alpha_\tau + \beta_\tau X_{t-1}, \tag{15}$$

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

$$U_{t,\tau} := Y_t - Q_\tau(Y_t \mid \mathcal{F}_{t-1}), \tag{16}$$

and substituting (15) into (16) yields $U_{t,\tau} = Y_t - \alpha_\tau - \beta_\tau X_{t-1}$. 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. Similar to (3) of the mean model, the object of interest here is to test :

$$H_0 : \beta_\tau = \beta_{\tau,0}$$

for some τ, where $\beta_{\tau,0}$ is a fixed hypothesized value. As before, in many applications, the null hypothesis of interest is often no quantile predictability, where is obtained by setting $\beta_{\tau,0} = 0$ so that $H_0 : \beta_\tau = 0$.

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 (15), 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 7. By definition, the quantile innovation $U_{t,\tau}$ in (16) satisfies $\mathbb{P}(U_{t,\tau} < 0 | \mathcal{F}_{t-1}) = \tau$. Hence, $\psi_\tau(U_{t,\tau}) = \tau - 1\{U_{t,\tau} < 0\}$ obeys

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

so that $\psi_\tau(U_{t,\tau})$ forms a martingale difference sequence with respect to $\mathcal{F}_{t-1}$ by construction. Therefore, we can show that the EL constraints remain valid, and the EL procedure developed for mean regressions extends naturally to the quantile setting. We note that $U_{t,\tau}$ itself need not be a martingale difference, and hence the conditions on U_t in Assumption 1 is no longer relevant in the quantile-EL context. Therefore, we instead assume the following:

Assumption 1′. *The predictor innovation ε_t is a martingale difference with respect to $\{\mathcal{F}_t\}$, $\sup_t \mathbb{E}|\varepsilon_t|^{2+q} < \infty$ for some $q > 0$, $\mathrm{Var}(\varepsilon_t) > 0$, and $\{(U_{t,\tau}, \varepsilon_t)\}$ is jointly stationary and ergodic.*

Theorem 4. *Suppose the data is generated according to the process in equations (15) and (2), and the predictive variable belongs to either class C1, C2, C3, C4, or C5. Suppose also Assumption 1′ holds. Then, for each $\tau \in (0,1)$, under $H_0 : \beta_\tau = \beta_{\tau,0}$ we have*

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

as $T \to \infty$, where $\beta_{\tau,0}$ is a fixed hypothesized value.

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 $\epsilon > 0$ such that for all t, the conditional density $f_{t,\tau}(0|\mathcal{F}_{t-1}) = f_\tau(0) \in (0,\infty)$, we have $|f_{t,\tau}(u \mid \mathcal{F}_{t-1}) - f_{t,\tau}(0|\mathcal{F}_{t-1})| \leq L_\tau|u|$ for all $|u| \leq \epsilon$.*

For the estimation of the unknown constant α_τ, we consider an extension of the approach of Demetrescu et al. (2025). For a given value of β_τ, we define

$$\widetilde{\alpha}_\tau(\beta_\tau) := \underset{a\in\mathbb{R}}{\operatorname{argmin}} \sum_{t=1}^{T} \vartheta_\tau \left(Y_t - a - \beta_\tau X_{t-1}\right). \tag{17}$$

Under $H_0 : \beta_\tau = \beta_{\tau,0}$, $\widetilde{\alpha}_\tau(\beta_{\tau,0})$ is the sample τ-quantile of $\{Y_t - \beta_{\tau,0}X_{t-1}\}_{t=1}^T$. As we show in the Appendix, under Assumptions 1′ and 3, $\widetilde{\alpha}_\tau(\beta_{\tau,0}) - \alpha_\tau = O_p(T^{-1/2})$ under all cases C1-C5.

We 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 (11), $\widetilde{\alpha}_\tau := \widetilde{\alpha}_\tau(\beta_\tau)$ 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

$$\sum_{t=1}^{T} \frac{\widetilde{\xi}_{t,\tau}(\beta_\tau)}{1+\lambda_\tau\widetilde{\xi}_{t,\tau}(\beta_\tau)} = 0. \tag{18}$$

Theorem 5. *Suppose the data is generated according to the process in equations (15) and (2) with $\theta = 0$, and the predictive variable belongs to either class C1, C2, C3, C4, or C5. Suppose also that Assumptions 1′ and 3 holds. Then, for each $\tau \in (0,1)$, under $H_0 : \beta_\tau = \beta_{\tau,0}$*

$$\widetilde{\ell}_T(\beta_{\tau,0}) \xrightarrow{d} \begin{cases} \chi^2_1 & \text{under C1–C2,} \\ \left(\sqrt{1-\lambda_\tau^2}\,Z + \lambda_\tau\mathcal{K}\right)^2 & \text{under C3-C5.} \end{cases}$$

as $T\to\infty$, where $\beta_{\tau,0}$ is a fixed hypothesized value, $\lambda_\tau := \frac{\mathbb{E}[\psi_\tau(U_{t,\tau})\varepsilon_t]}{\sigma_\varepsilon\sqrt{\tau(1-\tau)}}$, and $Z \sim N(0,1)$ is indpendent of $\mathcal{K}$, the persistence-specific random functional defined in Theorem 3.

Corollary 2. *In addition to what is assumed in Theorem 5, if* $\mathbb{E}[\psi_\tau(U_{t,\tau})\varepsilon_t] = 0$*, then*

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

as $T \to \infty$*, under all persistence classes C1, C2, C3, C4, and C5.*

The results above can be used to test the hypothesis $H_0 : \beta_\tau = \beta_{\tau,0}$. Furthermore, an empirical likelihood confidence set for β_τ can be obtained.

Remark 8. Theorem 5 and Corollary 2 establish Wilk's law and explains how contemporaneous score-innovation dependence affects the limiting distribution of the two-stage quantile EL statistic when the intercept is unknown.The condition $\mathbb{E}[\psi_\tau(U_{t,\tau})\varepsilon_t] = 0$ is weaker than contemporaneous independence between $U_{t,\tau}$ and ε_t. It eliminates the nonstandard component under C3-C5 and yields the chi-squared limit, whereas this restriction is not required under C1-C2. The role of score-innovation covariance is analogous to that in the nonstandard asymptotics of IVX predictive quantile regression analyzed by Lee (2016), although the Brownian functionals differ. We note that the functionals $\mathcal{K}$ are identical in both mean and quantile cases; details of their derivation can be found in the appendix.

In the simulation design in Section 5, we have $U_t = \phi\varepsilon_t + z_t$ with independent standard normal innovations and $U_{t,\tau} = U_t - \sqrt{1+\phi^2}\,\Phi^{-1}(\tau)$, and hence

$$\mathbb{E}[\psi_\tau(U_{t,\tau})\varepsilon_t] = \frac{\phi}{\sqrt{1+\phi^2}}\,\varphi\big(\Phi^{-1}(\tau)\big)\,,$$

and

$$\lambda_\tau = \frac{\phi}{\sqrt{1+\phi^2}}\,\frac{\varphi\big(\Phi^{-1}(\tau)\big)}{\sqrt{\tau(1-\tau)}},$$

where Φ and φ denote the standard normal distribution and density functions, respectively. Thus, the zero-covariance condition is automatically satisfied when $\phi = 0$, whereas nonzero ϕ deviates the limiting distribution under C3-C5. For each fixed τ, the limiting law approaches χ^2_1 as $\phi \to 0$.

This nonstandard component provides a theoretical mechanism consistent with the over-rejection observed under strong predictor persistence in Section 5. The magnitude and direction of the distortion depend on both λ_τ and the persistence-specific functional $\mathcal{K}$.

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 (14) (equivalently, in the quantile case, (18)) 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

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

By Taylor expansion to $\log(1+\widetilde{\lambda}\widetilde{Z}_t)$ in (13), it follows that

$$\begin{aligned}\widetilde{\ell}_T(\beta_0) &= 2\widetilde{\lambda}\sum_{t=1}^{T}\widetilde{Z}_t - \widetilde{\lambda}^2\sum_{t=1}^{T}\widetilde{Z}_t^2 + \frac{2\widetilde{\lambda}^3}{3}\sum_{t=1}^{T}\widetilde{Z}_t^3 - \frac{\widetilde{\lambda}^4}{2}\sum_{t=1}^{T}\widetilde{Z}_t^4 + \cdots \\ &= \frac{\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{Z}_t\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}_t^2} + f_T(\zeta) + o_p(1), \end{aligned} \tag{19}$$

where $f_T(\zeta)$ represents the influence of higher-order terms involving the third and fourth moments. Equation (19) 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

$$q = \frac{\text{kurtosis}(\widetilde{Z}_t)}{2} - \frac{\text{skewness}(\widetilde{Z}_t)^2}{3}. \tag{20}$$

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 $\text{mean}(|w(X_{t-1})| > 0.95)$. 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

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

q/T. Consequently, the test statistic (19) 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 9. 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 $U_t = \phi\varepsilon_t + z_t$; an additional component driven by ε_t appears, introduces additional dependence between the EL score and the persistent regressor, amplifying size distortion. See Remark 11 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:

$$\widehat{q} := \frac{\widehat{\text{kurtosis}}(\widetilde{Z}_t)}{2} - \frac{\widehat{\text{skewness}}(\widetilde{Z}_t)^2}{3} \quad \text{and} \quad \widetilde{\ell}_T(\beta_0)^{\text{naive}} := \frac{\widetilde{\ell}_T(\beta_0)}{1+\widehat{q}/T}. \tag{21}$$

Here, we need to emphasize that the correction is a heuristic; our simulations (See Section 5) indicate that (21) 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.

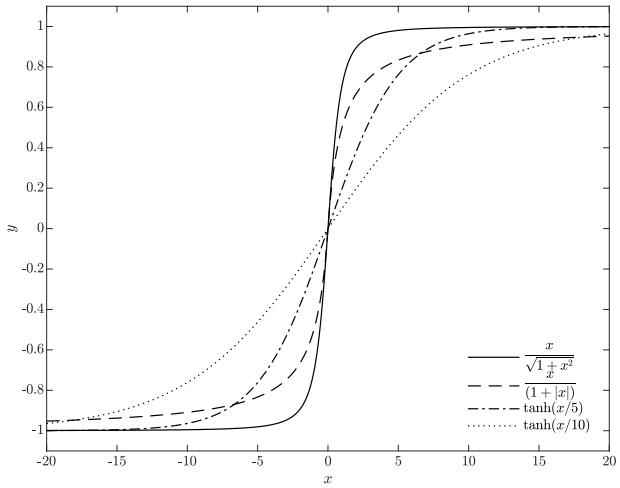

Figure 1. Plots of three weight functions

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 (8). The hyperbolic tangent weights $w_3(\cdot)$ approach ± 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 10. The choice of the scale b in $w(x) = \tanh(x/b)$ is an interesting problem in

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 $T = 250$.

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 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 $\alpha = 0.2$. We consider $c \in \{-50, -20, 0, 2\}$ and $a = \{0.9, 1\}$ to accommodate the cases of stationary, local-to-unity, unit root, and mildly explosive predictors. As done in Kostakis et al. (2023), we intialize the autoregressive process in equation (2) $X_0 = 0$, which is covered by the $X_0 = O_p(1)$ condition we impose. As a sensitivity check, we also include simulation result with $X_0 \sim N(0, 1)$ independently of the innovation, and obtain virtually identical results. This is expected because an $O_p(1)$ initial condition is asymptotically negligible relative to the stochastic scale of the predictor in C2-C5 and is geometrically attenuated in C1, see Phillips and Magdalinos (2007). The vector $(z_t', \varepsilon_t')'$ is drawn from a bivariate standard normal distribution, and we set $U_t = \phi\varepsilon_t + z_t$ as defined in (4), with $\phi \in \{-0.5, -0.2, 0\}$.[1] For quantile regressions, we use the same simulation setup for X_t, ε_t, and U_t, but generate Y_t from

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

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

[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.

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.

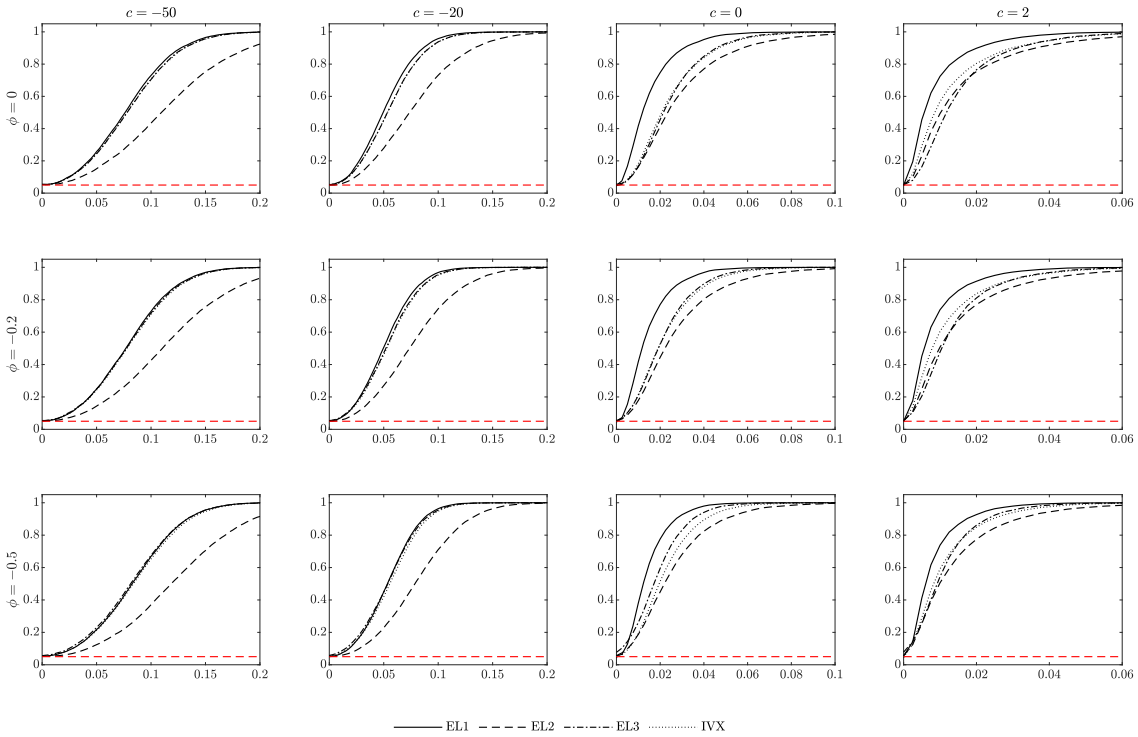

Figure 2. Finite-sample size and power plots for mean predictability tests with homoscedastic errors and tanh-based weight

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.

Remark 11. 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 9. When $\phi \neq 0$, the EL score in (12) inherits the influence of the term $\phi\varepsilon_t$, 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 $\phi\varepsilon_t$ 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.

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.

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

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 (15), 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 $T = 250$.

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 ($c = 0$),

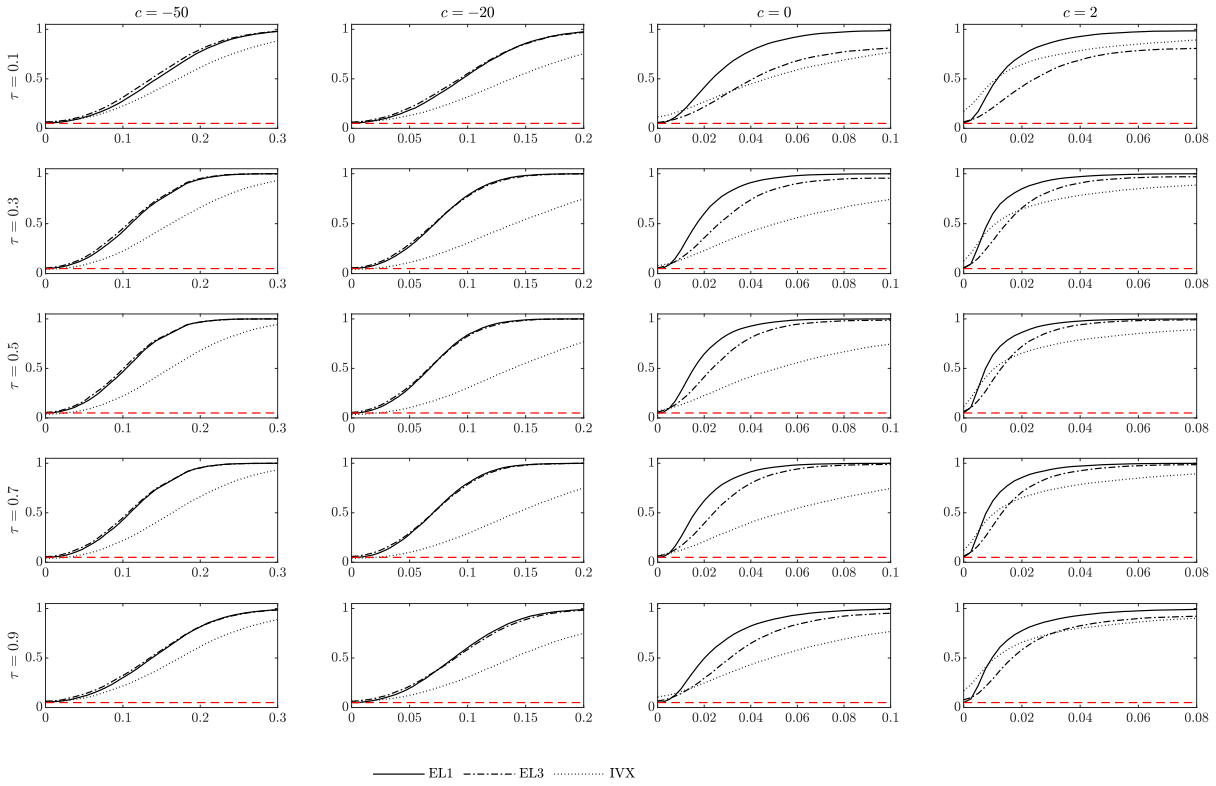

Figure 3. Finite-sample size and power plots for quantile predictability tests with homoscedastic errors and correlation $\phi = -0.5$

Note: The figure summarizes rejection probabilities (y-axis) for tests of the null hypothesis of no predictive ability in the quantile regression model (15), 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 $\phi = -0.5$.

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 12. We note that the mild over-rejection observed for mean predictability under strong endogeneity (cf. Remark 11) 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 $\phi\varepsilon_t$ affects 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 $U_t = \phi\varepsilon_t + z_t$. 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.

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

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

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.

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 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

[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.

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.

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

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., M. M. Rodrigues, Paulo, and A. M. Taylor, Robert (2025). Predictive quantile regressions with persistent and heteroskedastic predictors: A powerful 2sls testing approach. *Journal of Econometrics 249*, 106002.

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.

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.

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

Kostakis, A., T. Magdalinos, and M. P. Stamatogiannis (2023). Taking stock of long-horizon predictability tests: Are factor returns predictable? *Journal of Econometrics 237*(1), 105–380.

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. (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.

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

Throughout, we use C and C' to denote some generic constants whose value may vary between occurrences, unless defined otherwise. The initial condition $X_0 = O_p(1)$.

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 desired limit via the martingale central limit theorem and a common-sign approximation. Write $D_{Tt} = X_{m^*+t-1} - X_{t-1}, U_t^* = U_{m^*+t} - U_t, Z_{Tt}^* = U_t^* w(D_{Tt})$, and put $S_T = \sum_{t=1}^m Z_{Tt}^*$ and $Q_T = \sum_{t=1}^m (Z_{Tt}^*)^2$. All quantities below are evaluated under $H_0 : \beta = \beta_0$. Furthermore, as discussed in the paper, $|w| \le 1$, w is globally Lipschitz,

$$|w(x) - \mathrm{sgn}(x)| \le C|x|^{-1}, \qquad |x| \ge 1. \tag{22}$$

Put $\Gamma_t = (z_t, \varepsilon_t)'$ and $\mathcal{F}_t = \sigma(\Gamma_s : s \le t)$, and write $h_t = \mathbb{E}(U_t^2 \mid \mathcal{F}_{t-1})$. By the tower property, U_t and ε_t are martingale differences with respect to $\mathcal{F}_t$, and $h_t \ge c$ a.s. Moreover, h_t is stationary and ergodic, with $\mathbb{E}h_t = \sigma^2$. We write $\sigma_\varepsilon^2 = \mathbb{E}(\varepsilon_0^2) > 0$ and put $p = \min(2, 1 + q/2) > 1$.

We first study the stationary case C1. Let $X_t' = \sum_{j=0}^\infty \rho^j \varepsilon_{t-j}$ be the stationary solution, and put $Z_{Tt}' = (U_{m^*+t} - U_t) w(X_{m^*+t-1}' - X_{t-1}')$. Since $X_t - X_t' = \rho^t (X_0 - X_0')$, Lipschitz continuity gives

$$\sum_{t=1}^m |Z_{Tt}^* - Z_{Tt}'| \;\le\; C|X_0 - X_0'| \sum_{t=1}^m |\rho|^{t-1} |U_t^*| = O_p(1).$$

The same argument, with $(U_t^*)^2$ in place of $|U_t^*|$, shows that

$$\frac{1}{\sqrt{m}} \sum_{t=1}^m (Z_{Tt}^* - Z_{Tt}') = o_p(1), \qquad \frac{1}{m} \sum_{t=1}^m \{(Z_{Tt}^*)^2 - (Z_{Tt}')^2\} = o_p(1). \tag{23}$$

Only tightness of $X_0 - X_0'$ is required here.

We next separate the two blocks by a finite discretization. Put $r = \lfloor T/8 \rfloor$ and $X_s'^{[r]} = \sum_{j=0}^{r-1} \rho^j \varepsilon_{s-j}$, and consider the $2m$-dimensional vector $B_T = \big(U_{m^*+t}, X_{m^*+t-1}'^{[r]} : 1 \le t \le m\big)$. Its generating innovations are separated from $\mathcal{F}_m$ by $g_T = m^* - r + 1 - m \ge T/4$ for all sufficiently large T.

Partition $[-m, m]$ into intervals of length $\delta_m = m^{-3}$, add the two tail intervals, and apply the resulting quantizer coordinatewise to B_T. Denote the resulting finite-valued vector by A_T. The number N_T of its possible values satisfies $N_T \le (2m^4 + 3)^{2m}$. For every atom a of A_T, $\mathbb{E}\,|\mathbb{P}(A_T = a|\mathcal{F}_m) - \mathbb{P}(A_T = a)| \le 2\alpha_{\mathrm{mix}}(g_T)$.

Consequently, the expected total variation distance between its conditional and marginal laws is bounded by

$$N_T \alpha_{\mathrm{mix}}(g_T) \le C_0 \exp\{8m \log m + O(m) - c_0 T/4\} = o(1). \tag{24}$$

Finite-state maximal coupling therefore gives a copy $A_T^\dagger$, independent of $\mathcal{F}_m$, with $\mathbb{P}(A_T^\dagger \neq A_T) = o(1)$. Using fresh auxiliary randomness and the original regular conditional law of the entire error process given A_T, extend $A_T^\dagger$ to a two-sided copy $\{\Gamma_s^\dagger\}$, independent of $\mathcal{F}_m$ and having the original stationary law. Define $U_s^\dagger$, $X_s'^\dagger$ and $X_s^{\prime[r],\dagger}$ accordingly.

The $2+q$ moments of U_s, X_s' and $X_s^{\prime[r]}$ are bounded uniformly in s and r. Hence, by the union bound, all $O(m)$ original and copied errors, truncated predictors, and predictor-tail variables needed for the following comparison have absolute value at most m with probability tending to one. On this event and $\{A_T^\dagger = A_T\}$,

$$\max_{t\le m}|U_{m^*+t} - U_{m^*+t}^\dagger| \le \delta_m, \qquad \max_{t\le m}|X'_{m^*+t-1} - X'^\dagger_{m^*+t-1}| \le \delta_m + 2m|\rho|^r,$$

because $X_s' - X_s^{\prime[r]} = \rho^r X'_{s-r}$. Writing $\check{Z}_{Tt}^* = (U_{m^*+t}^\dagger - U_t)w(X'^\dagger_{m^*+t-1} - X'_{t-1})$, we obtain $\max_{t\le m}|Z'_{Tt} - \check{Z}_{Tt}^*| \le C(m^{-2} + m^2|\rho|^r)$ on the same event. It follows from (23) that

$$\frac{S_T - \sum_{t=1}^m \check{Z}_{Tt}^*}{\sqrt{m}} = o_p(1), \qquad \frac{Q_T - \sum_{t=1}^m (\check{Z}_{Tt}^*)^2}{m} = o_p(1). \tag{25}$$

Indeed, the two discretization errors are bounded by $C(m^{-3/2} + m^{5/2}|\rho|^r)$ and $C(m^{-1} + m^3|\rho|^r)$, respectively. Let $\mathcal{F}_s^\dagger = \sigma(\Gamma_j^\dagger : j \le s)$ and $\mathcal{H}_{T,t} = \mathcal{F}_t \vee \mathcal{F}_{m^*+t}^\dagger$, $0 \le t \le m$.

Since the copied process is independent of $\mathcal{F}_m$, $\{\check{Z}_{Tt}^*\}$ is a martingale difference array with respect to $\mathcal{H}_{T,t}$, and

$$\mathbb{E}[(\check{Z}_{Tt}^*)^2 \mid \mathcal{H}_{T,t-1}] = (h_t + h_{m^*+t}^\dagger)w(X'^\dagger_{m^*+t-1} - X'_{t-1})^2. \tag{26}$$

Two independent copies of the strongly mixing error process form an ergodic process. The sample in (26) has the law of a sample from these two independent stationary copies. Therefore,

$$\frac{1}{m}\sum_{t=1}^m \mathbb{E}[(\check{Z}_{Tt}^*)^2 \mid \mathcal{H}_{T,t-1}] \xrightarrow{p} \eta_\rho^2 := \mathbb{E}\left[(h_1^{(1)} + h_1^{(2)})w(X_0'^{(2)} - X_0'^{(1)})^2\right] > 0. \tag{27}$$

Strict positivity follows from $h_t \ge c$, nondegeneracy of X_0', and strict monotonicity of the three weights. In particular, conditional heteroscedasticity does not require η_ρ^2 to factor into two expectations. For each $\epsilon > 0$, boundedness of w and the moment condition give

$$\mathbb{E}\left[\frac{1}{m}\sum_{t=1}^m \mathbb{E}\left\{(\check{Z}_{Tt}^*)^2 1\{|\check{Z}_{Tt}^*| > \epsilon\sqrt{m}\}|\mathcal{H}_{T,t-1}\right\}\right] \le C_\epsilon m^{-q/2} \longrightarrow 0.$$

Thus the conditional Lindeberg condition holds in probability. The martingale L^p inequality also gives

$$\mathbb{E}\left|\frac{1}{m}\sum_{t=1}^m \left\{(\check{Z}_{Tt}^*)^2 - \mathbb{E}[(\check{Z}_{Tt}^*)^2 \mid \mathcal{H}_{T,t-1}]\right\}\right|^p \le Cm^{1-p} \longrightarrow 0.$$

By the martingale central limit theorem of Hall and Heyde (1980) and (25),

$$S_T/\sqrt{m} \xrightarrow{d} N(0, \eta_\rho^2), \qquad Q_T/m \xrightarrow{p} \eta_\rho^2. \tag{28}$$

Now we move on to the nonstationary cases C2-C5. No mixing condition is necessary in the remainder of this proof. First, we write the unweighted sum in chronological order as

$$\sum_{t=1}^{m} U_t^* = -\sum_{t=1}^{m} U_t + \sum_{t=m^*+1}^{m^*+m} U_t.$$

This is a martingale array with deterministic coefficients. Its conditional variance, divided by m, converges to $2\sigma^2$, since $\frac{1}{m}\sum_{t=1}^{m} h_t \xrightarrow{p} \sigma^2$ and $\frac{1}{m}\sum_{t=1}^{m} h_{m^*+t} \xrightarrow{p} \sigma^2$.

We note that the second convergence follows by stationarity: the moving block average has the same distribution as the first average. The expected conditional Lindeberg sum is again $O(m^{-q/2})$. Moreover, both averages of squared errors converge to σ^2 by ergodicity, whereas $\{U_t U_{m^*+t}\}_{t\le m}$ is a martingale difference array with respect to $\mathcal{F}_{m^*+t}$. Since $2p \le 2+q$,

$$\mathbb{E}\left|\frac{1}{m}\sum_{t=1}^{m} U_t U_{m^*+t}\right|^p \le C m^{1-p} \longrightarrow 0.$$

Consequently,

$$\frac{\sum_{t=1}^{m} U_t^*}{\sqrt{2m\sigma^2}} \xrightarrow{d} N(0,1), \qquad \frac{1}{m}\sum_{t=1}^{m}(U_t^*)^2 \xrightarrow{p} 2\sigma^2. \tag{29}$$

Next, by (2),

$$D_{Tt} = \rho_T^{t-1} D_{T1} + \sum_{j=1}^{t-1} \rho_T^{t-1-j}(\varepsilon_{m^*+j} - \varepsilon_j), \qquad D_{T1} = X_{m^*} - X_0. \tag{30}$$

Expanding D_{T1} and applying Lemma 2.1(ii) of Abadir, Distaso, Giraitis, and Koul (2014) yields

$$D_{T1}/d_T \xrightarrow{d} N(0,\tau_c^2), \qquad (d_T, \tau_c^2) = \begin{cases} \left(T^{a/2},\, \sigma_\varepsilon^2/(-2c)\right), & \text{C2,} \\ \left(\sqrt{T},\, \sigma_\varepsilon^2 \displaystyle\int_0^{1/2} e^{2c(1/2-s)}\, ds\right), & \text{C3–C4,} \\ \left(\rho_T^{m^*} T^{a/2},\, \sigma_\varepsilon^2/(2c)\right), & \text{C5,} \end{cases} \tag{31}$$

where $c = 0$ in C4. Indeed, $D_{T1} = (\rho_T^{m^*} - 1)X_0 + \sum_{j=1}^{m^*} \rho_T^{m^*-j}\varepsilon_j$. The initial term divided by d_T is $o_p(1)$. The geometric coefficients, normalized to have squared sum one and extended by zero, satisfy $\sum_j (b_{Tj} - b_{T,j-1})^2 = o(1)$. The cited lemma applies to stationary ergodic martingale differences; summing the squared geometric coefficients gives the variances in (31).

Under C2, $m = o(T^a)$ and Doob's inequality applied separately to the early and late innovation sums give $\max_{t\le m} \frac{|D_{Tt} - D_{T1}|}{T^{a/2}} = O_p(m/T^a) + O_p(\sqrt{m/T^a}) = o_p(1)$. Under C3-C4, the same argument and $m = o(T)$ yield $\max_{t\le m} \frac{|D_{Tt} - D_{T1}|}{\sqrt{T}} = O_p(m/T) + O_p(\sqrt{m/T}) = o_p(1)$.

Under C5, discounting (30) gives $\max_{t\le m}|\rho_T^{-(t-1)}D_{Tt} - D_{T1}| = O_p(T^{a/2}) = o_p(d_T)$, since $\sum_{j\ge 1}\rho_T^{-2j} = O(T^a)$.

It follows from these bounds and (31) that, with $s_T = \operatorname{sgn}(D_{T1})$ (assigned either sign at zero),

$$\lim_{\delta\downarrow 0}\limsup_{T\to\infty}\mathbb{P}\left(\min_{t\le m} s_T D_{Tt} < \delta d_T\right) = 0. \tag{32}$$

For C5, this also uses $\rho_T^{t-1} \ge 1$.

On the complementary event in (32), (22) implies

$$\frac{|S_T - s_T\sum_{t=1}^m U_t^*|}{\sqrt{m}} \le \frac{C}{\delta d_T\sqrt{m}}\sum_{t=1}^m |U_t^*| = O_p\left(\frac{\sqrt{m}}{\delta d_T}\right) = o_p(1).$$

Here $\sqrt{m}/d_T \to 0$ under the stated pair-count restrictions. Similarly,

$$\left|\frac{Q_T}{m} - \frac{1}{m}\sum_{t=1}^m (U_t^*)^2\right| \le \frac{C}{\delta d_T m}\sum_{t=1}^m (U_t^*)^2 = o_p(1).$$

Letting $\delta \downarrow 0$ therefore gives

$$S_T = s_T\sum_{t=1}^m U_t^* + o_p(\sqrt{m}), \qquad Q_T/m \xrightarrow{p} 2\sigma^2. \tag{33}$$

Since $s_T^2 = 1$,(29) and (33) imply

$$S_T^2/Q_T \xrightarrow{d} \chi_1^2.$$

No independence between s_T and the unweighted error sum is required.

We now study the empirical likelihood expansion. In all cases, $S_T = O_p(\sqrt{m})$, Q_T/m converges to a strictly positive constant, and

$$\mathbb{P}\left(\max_{t\le m}|Z_{Tt}^*| > \epsilon\sqrt{m}\right) \le C_\epsilon m^{-q/2} \to 0.$$

These properties also imply the sample convex-hull condition with probability tending to one. Otherwise, all nonzero scores would have the same sign and $Q_T \le \max_t |Z_{Tt}^*|\,|S_T| = o_p(m)$, which is a contradiction.

Writing $M_T = \max_t |Z_{Tt}^*|$, the multiplier equation gives

$$|\lambda| \le \frac{|S_T|}{Q_T - M_T|S_T|} = O_p(m^{-1/2}), \qquad |\lambda|M_T = o_p(1),$$

on an event whose probability tends to one. Taylor expansion consequently yields

$$\lambda = \frac{S_T}{Q_T} + o_p(m^{-1/2})$$

and

$$\ell_T^*(\beta_0) = 2\lambda S_T - \lambda^2 Q_T + o_p(1) = \frac{S_T^2}{Q_T} + o_p(1) \xrightarrow{d} \chi_1^2,$$

which completes the proof. □

Proof of Theorem 3. We write $\widetilde{Y}_t = Y_t - \widetilde{\alpha}$ and $\widetilde{Z}_t(\beta_0) = [\widetilde{Y}_t - \beta_0 X_{t-1}]w^c(X_{t-1})$, recall that the centered weight is $w^c(X_{t-1}) := w(X_{t-1}) - \frac{1}{T}\sum_{s=1}^T w(X_{s-1})$. We have

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^T \widetilde{Z}_t(\beta_0) &= \frac{1}{\sqrt{T}}\sum_{t=1}^T \big[(Y_t - \alpha - \beta_0 X_{t-1}) - (\widetilde{\alpha} - \alpha)\big] w^c(X_{t-1}) \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^T U_t w^c(X_{t-1}) - \sqrt{T}(\widetilde{\alpha} - \alpha)\frac{1}{T}\sum_{t=1}^T w^c(X_{t-1}) =: A_T - B_T, \end{aligned} \tag{34}$$

and $B_T \equiv 0$ by the construction of the centered weight.

It remains to establish the limiting distribution of A_T after studentization, consistency of the corresponding second-moment normalizer, and the negligibility of the largest estimating function. We write $w_t := w(X_{t-1})$, $\overline{w}_T := T^{-1}\sum_{s=1}^T w_s$, $w_t^c := w_t - \overline{w}_T$, $J_T := \sum_{t=1}^T (w_t^c)^2$.

Since U_t is a martingale difference with respect to $\mathcal{G}_{t-1} = \sigma(X_0, \{\varepsilon_s, z_s : s \le t-1\})$ and $w_t = w(X_{t-1})$ is $\mathcal{G}_{t-1}$-measurable, the vector $\Delta_{Tt} := (U_t w_t \;\; U_t)'$ is a martingale difference array. Consider

$$M_T := \frac{1}{\sqrt{T}}\sum_{t=1}^T \Delta_{Tt} = \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T U_t w_t, \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T U_t \end{pmatrix} =: \begin{pmatrix} A_{1T} \\ A_{0T} \end{pmatrix}, \tag{35}$$

so that $A_T = A_{1T} - \overline{w}_T A_{0T}$. As in Assumption 2, let $\sigma_t^2 := \mathbb{E}(U_t^2|\mathcal{G}_{t-1})$, which equals σ^2 under conditional homoscedasticity and its sample average limits to σ^2 under conditional heteroscedasticity. Its predictable quadratic variation matrix is

$$\langle M\rangle_T := \frac{1}{T}\sum_{t=1}^T \mathbb{E}(\Delta_{Tt}\Delta_{Tt}'|\mathcal{G}_{t-1}) = \frac{1}{T}\sum_{t=1}^T \sigma_t^2 \begin{pmatrix} w_t^2 & w_t \\ w_t & 1 \end{pmatrix}. \tag{36}$$

Accordingly, define the quadratic form

$$\eta_T^2 := \begin{pmatrix} 1 & -\overline{w}_T \end{pmatrix} \langle M\rangle_T \begin{pmatrix} 1 \\ -\overline{w}_T \end{pmatrix} = \frac{1}{T}\sum_{t=1}^T \sigma_t^2 (w_t^c)^2. \tag{37}$$

Note that $U_t w_t^c$ is not a martingale difference since $\overline{w}_T$ is not generally $\mathcal{G}_{t-1}$-measurable; thus, a martingale central limit theorem cannot be directly applied. We hence derive the joint limits separately across persistence classes rather than infer a mixed-normal limit directly from (37).

Ultimately, we aim to establish the following under each class C1-C5:

$$\frac{A_T}{\sqrt{\eta_T^2}} = \frac{\sum_{t=1}^T U_t w_t^c}{\sqrt{\sum_{t=1}^T \sigma_t^2 (w_t^c)^2}} \Rightarrow \text{some distributional limit} \tag{38}$$

$$\frac{V_T}{\eta_T^2} = \frac{\sum_{t=1}^T U_t^2 (w_t^c)^2}{\sum_{t=1}^T \sigma_t^2 (w_t^c)^2} \xrightarrow{p} 1 \tag{39}$$

(where $\sum_{t=1}^T \sigma_t^2 (w_t^c)^2$ simplifies to $\sigma^2 J_T$ in the conditional homoscedastic case), after which the desired result follows, since we later show that $\widetilde{\ell}_T(\beta_0) = 2\sum_{t=1}^T \log\{1 + \widetilde{\lambda}\widetilde{Z}_t(\beta_0)\} = \{\sum_t \widetilde{Z}_t(\beta_0)\}^2 / \sum_t \widetilde{Z}_t(\beta_0)^2 + o_p(1) = A_T^2/V_T + o_p(1)$.

Write $\sigma_\varepsilon^2 := \mathbb{E}(\varepsilon_t^2) > 0$, and put $p = 1 + \min\{q/2, 1\} \in (1, 2]$. Conditional Jensen's inequality and Assumptions 1 and 2 give $\sup_t \mathbb{E}[|U_t^2 - \sigma_t^2|^p + (\sigma_t^2)^p] < \infty$. Since $\{U_t^2 - \sigma_t^2\}$ is a martingale difference, the martingale L^p inequality and Doob's maximal inequality imply that for every fixed $R \in (0, \infty)$ and deterministic $n \to \infty$,

$$\mathbb{E}\left[\max_{0 \le m \le \lfloor nR \rfloor} \left| \frac{1}{n} \sum_{t=1}^m (U_t^2 - \sigma_t^2) \right|^p \right] \le \frac{C}{n^p} \sum_{t=1}^{\lfloor nR \rfloor} \mathbb{E}|U_t^2 - \sigma_t^2|^p \le C' n^{1-p} = o(1). \tag{40}$$

Note that we can take $n = T^a$ for C5, and otherwise $n = T$.

Moreover, stationarity and ergodicity of $\{U_t^2\}$ imply that its normalized partial sums converge uniformly on compact intervals, almost surely, to $r\sigma^2$. Combining these results we obtain

$$\sup_{0 \le r \le R} \left| \frac{1}{n} \sum_{t=1}^{\lfloor nr \rfloor} U_t^2 - \sigma^2 r \right| + \sup_{0 \le r \le R} \left| \frac{1}{n} \sum_{t=1}^{\lfloor nr \rfloor} \sigma_t^2 - \sigma^2 r \right| \xrightarrow{p} 0, \tag{41}$$

which proves (6). By stationarity and ergodicity of (U_t, ε_t), as derived in Lee (2016), the joint martingale functional central limit theorem gives

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \begin{pmatrix} U_t \\ \varepsilon_t \end{pmatrix} \Longrightarrow \begin{pmatrix} \sigma B_U(r) \\ \sigma_\varepsilon B_\varepsilon(r) \end{pmatrix}. \tag{42}$$

Indeed, the normalized quadratic covariations converge to their unconditional second moments by ergodicity, and the $2 + q$ moment condition makes the normalized maximal jumps negligible, Hall and Heyde (1980). Here B_U and B_ε are standard Brownian motions with $\mathbb{E}[B_U(r) B_\varepsilon(r)] = \lambda_U r$ and $\lambda_U := \frac{\mathbb{E}[U_t \varepsilon_t]}{\sigma \sigma_\varepsilon}$. On an enlarged probability space, we may write

$$B_U(r) = \sqrt{1 - \lambda_U^2}\, B_\perp(r) + \lambda_U B_\varepsilon(r), \tag{43}$$

where $B_\perp$ is a standard Brownian motion independent of B_ε. We have

$$V_T := \frac{1}{T}\sum_{t=1}^{T}(U_t w_t^c)^2 = \frac{1}{T}\sum_{t=1}^{T}U_t^2 w_t^2 - 2\overline{w}_T\frac{1}{T}\sum_{t=1}^{T}U_t^2 w_t + \overline{w}_T^2\frac{1}{T}\sum_{t=1}^{T}U_t^2$$
$$=: A_T^{(2)} + B_T^{(2)} + C_T^{(2)}. \tag{44}$$

Similarly, define $A_T^{(2,0)} := T^{-1}\sum_{t=1}^{T}\sigma_t^2 w_t^2$, $B_T^{(2,0)} := T^{-1}\sum_{t=1}^{T}\sigma_t^2 w_t$, $C_T^{(2,0)} := T^{-1}\sum_{t=1}^{T}\sigma_t^2$. For $k = 0, 1, 2$, $(U_t^2 - \sigma_t^2)w_t^k$ is a martingale difference. The moment assumption and boundedness of w imply

$$\frac{1}{T}\sum_{t=1}^{T}(U_t^2 - \sigma_t^2)w_t^k = o_p(1). \tag{45}$$

Therefore, $A_T^{(2)} - A_T^{(2,0)} = o_p(1)$, $B_T^{(2)} + 2\overline{w}_T B_T^{(2,0)} = o_p(1)$, and $C_T^{(2)} - \overline{w}_T^2 C_T^{(2,0)} = o_p(1)$, so that

$$V_T = A_T^{(2,0)} - 2\overline{w}_T B_T^{(2,0)} + \overline{w}_T^2 C_T^{(2,0)} + o_p(1)$$
$$= \frac{1}{T}\sum_{t=1}^{T}\sigma_t^2(w_t^c)^2 + o_p(1)$$
$$= \sigma^2\eta_T^2 + o_p(1). \tag{46}$$

We note that a sufficient condition for (46) to imply the required statement (39), i.e. $V_T/\eta_T^2 \to^p 1$, is $1/\eta_T^2 = O_p(1)$. Under C1-C4, below we show either that $\eta_T^2 \to^p c > 0$ or $\eta_T^2 \Rightarrow \xi$ where ξ is strictly positive a.s., so that $1/\eta_T^2 = O_p(1)$ and (39) follows. For C5, we have $J_T/T = o_p(1)$ instead, which implies $\eta_T^2 = o_p(1)$ (via Hölder's inequality in the conditional heteroscedastic case; trivial in the conditional homoscedastic case because

$$\eta_T^2 = \frac{1}{T}\sum_{t=1}^{T}\sigma_t^2(w_t^c)^2 = \sigma^2\frac{1}{T}\sum_{t=1}^{T}(w_t^c)^2 \equiv \frac{J_T}{T} \tag{47}$$

with $\sigma_t^2 = \sigma^2$), and thus we shall establish the stronger relative approximation later.

Case 1: C1. Let $X_t' := \sum_{j=0}^{\infty}\rho^j\varepsilon_{t-j}$ be the stationary solution. Since $X_t - X_t' = \rho^t(X_0 - X_0')$ and $|\rho| < 1$, boundedness and uniform continuity of w, together with ergodicity, give $\overline{w}_T \to^p \mathbb{E}[w(X_0')]$, $J_T/T \to^p \mathrm{Var}\{w(X_0')\} > 0$. Truncating U_t^2 and using the same stationary approximation, followed by (45), yields

$$\eta_T^2 \xrightarrow{p} \eta^2 := \mathbb{E}\big[U_1^2\{w(X_0') - \mathbb{E}[w(X_0')]\}^2\big] > 0. \tag{48}$$

The strict positivity in (48) follows by Assumption 2. Under conditional homosecdasticity of U, it reduces to $\sigma_U^2\,\mathrm{Var}\{w(X_0')\} > 0$.

The weights $w_t - \mathbb{E}[w(X_0')]$ are predictable. Their conditional variance average converges to η^2 and the following verifies the Lindeberg condition. Given that w is uniformly bounded,

$\|\Delta_{Tt}\| \le C|U_t|$. Using $\sup_t \mathbb{E}|U_t|^{2+q} < \infty$, for every $\epsilon_0 > 0$,

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\Big[\|\Delta_{Tt}\|^2 1\{\|\Delta_{Tt}\| > \epsilon_0\sqrt{T}\} \mid \mathcal{G}_{t-1}\Big] \le \frac{C}{T^{q/2}}\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(|U_t|^{2+q} \mid \mathcal{G}_{t-1}) = o_p(1). \tag{49}$$

For later reference, we note that the same argument applies with T^a in place of T over $t \le \lfloor T^a R \rfloor$, for every fixed $R < \infty$. It now follows that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_t\{w_t - \mathbb{E}[w(X_0')]\} \Rightarrow N(0, \eta^2). \tag{50}$$

Since $A_{0T} = O_p(1)$, we have $(\overline{w}_T - \mathbb{E}[w(X_0')])A_{0T} = o_p(1)$. Also, $A_T \Rightarrow N(0, \eta^2)$ and, by (46) and (48), $V_T \xrightarrow{p} \eta^2$. Therefore

$$\frac{A_T}{\sqrt{V_T}} \Rightarrow N(0, 1). \tag{51}$$

Case 2: C2. Let $\rho_T = 1 + c/T^a$ with $c < 0$ and $0 < a < 1$. The same argument for the predictor and weights in the proof of Theorem 5 gives $\overline{w}_T = o_p(1)$, $\frac{1}{T}\sum_{t=1}^{T} w_t^2 \to^p 1$, $J_T/T \to^p 1$. These conclusions involve only the predictor and the weight function, so the same argument applies here. Uniform integrability of σ_t^2 and boundedness of w imply

$$\frac{1}{T}\sum_{t=1}^{T} \sigma_t^2(w_t^2 - 1) = o_p(1). \tag{52}$$

To see this, we truncate σ_t^2 at a fixed level, apply $T^{-1}\sum_t |w_t^2 - 1| = o_p(1)$ to the bounded part, and let the truncation level tend to infinity. Consequently, $\frac{1}{T}\sum_{t=1}^{T} \sigma_t^2 w_t^2 \xrightarrow{p} \sigma^2$, and $\eta_T^2 \xrightarrow{p} \sigma^2 > 0$. The martingale CLT, Hall and Heyde (1980) and (49) yield $A_{1T} \Rightarrow N(0, \sigma^2)$.

Since $\overline{w}_T A_{0T} = o_p(1)$, we have $A_T \Rightarrow N(0, \sigma^2)$. Equation (46) gives $V_T \xrightarrow{p} \sigma^2$, and therefore

$$\frac{A_T}{\sqrt{V_T}} \Rightarrow N(0, 1). \tag{53}$$

Case 3: C3. Let $\rho_T = 1 + c/T$ with $c \neq 0$. By (42) and the AR recursion,

$$\frac{X_{\lfloor Tr \rfloor}}{\sigma_\varepsilon\sqrt{T}} \Longrightarrow \mathcal{J}_c(r) := \int_0^r e^{c(r-s)}\, dB_\varepsilon(s). \tag{54}$$

The weight argument in the proof of Theorem 5 gives, jointly,

$$\overline{w}_T \Rightarrow \int_0^1 \operatorname{sgn}\{\mathcal{J}_c(r)\}\, dr, \qquad \frac{J_T}{T} \Rightarrow \mathcal{V}_c := 1 - \left[\int_0^1 \operatorname{sgn}\{\mathcal{J}_c(r)\}\, dr\right]^2. \tag{55}$$

Here $\mathcal{V}_c > 0$ almost surely, because $\mathcal{J}_c$ has zero occupation time at zero and spends positive time on both half-lines.

For a fixed $\delta > 0$, approximate sgn by a Lipschitz function $g_\delta : \mathbb{R} \to [-1, 1]$ agreeing with it outside $[-\delta, \delta]$. For the smoothed predictable weights, stochastic-integral convergence follows

from Kurtz and Protter (1991): the normalized U_t martingales have uniformly bounded expected quadratic variation on compact intervals. The smoothing error vanishes by Itô isometry, saturation of w, and the zero occupation time at zero. Under conditional heteroscedasticity, we first truncate σ_t^2 in the isometry bound and then use its uniform integrability to remove the truncation. Thus, jointly with (55), in view of $A_T = A_{1T} - \overline{w}_T A_{0T}$ we have

$$\frac{A_T}{\sigma} \Rightarrow \int_0^1 \operatorname{sgn}\{\mathcal{J}_c(r)\}\, dB_U(r) - \left[\int_0^1 \operatorname{sgn}\{\mathcal{J}_c(r)\}\, dr\right] B_U(1). \tag{56}$$

For the variance, the partial-sum laws of large numbers give

$$\sup_{0\le r\le 1}\left|\frac{1}{T}\sum_{t=1}^{\lfloor Tr\rfloor}\sigma_t^2 - \sigma^2 r\right| \xrightarrow{p} 0, \qquad \sup_{0\le r\le 1}\left|\frac{1}{T}\sum_{t=1}^{\lfloor Tr\rfloor}U_t^2 - \sigma^2 r\right| \xrightarrow{p} 0. \tag{57}$$

We apply these laws to the smoothed weights and their squares, and then remove the smoothing by truncation and the zero-occupation-time argument. This yields

$$\frac{1}{T}\sum_{t=1}^{T}\sigma_t^2 w_t^k - \frac{\sigma^2}{T}\sum_{t=1}^{T} w_t^k = o_p(1), \qquad k = 0, 1, 2. \tag{58}$$

Consequently,

$$\eta_T^2 - \sigma^2\frac{J_T}{T} = o_p(1), \qquad V_T - \sigma^2\frac{J_T}{T} = o_p(1), \tag{59}$$

and $V_T/\eta_T^2 \xrightarrow{p} 1$. Combining (43), (56), and (59) gives

$$\frac{A_T}{\sqrt{V_T}} \Rightarrow \sqrt{1-\lambda_U^2}\, Z + \lambda_U \mathcal{K}_3, \tag{60}$$

where $Z \sim N(0,1)$ and

$$\mathcal{K}_3 := \mathcal{V}_c^{-1/2}\left[\int_0^1 \operatorname{sgn}\{\mathcal{J}_c(r)\}\, dB_\varepsilon(r) - \left(\int_0^1 \operatorname{sgn}\{\mathcal{J}_c(r)\}\, dr\right) B_\varepsilon(1)\right]. \tag{61}$$

Indeed, conditional on the entire path of B_ε, the corresponding $B_\perp$ expression divided by $\sqrt{\mathcal{V}_c}$ is $N(0,1)$, and its conditional law does not depend on B_ε. Hence Z is independent of $\mathcal{K}_3$.

Case 4: C4. Setting $c = 0$ in the preceding argument gives $\mathcal{J}_0(r) = B_\varepsilon(r)$ and

$$\frac{J_T}{T} \Rightarrow \mathcal{V}_0 := 1 - \left[\int_0^1 \operatorname{sgn}\{B_\varepsilon(r)\}\, dr\right]^2 > 0 \quad \text{a.s.} \tag{62}$$

Moreover, $V_T/\eta_T^2 \xrightarrow{p} 1$ and

$$\frac{A_T}{\sqrt{V_T}} \Rightarrow \sqrt{1-\lambda_U^2}\, Z + \lambda_U \mathcal{K}_4, \tag{63}$$

where

$$\mathcal{K}_4 := \mathcal{V}_0^{-1/2}\left[\int_0^1 \operatorname{sgn}\{B_\varepsilon(r)\}\, dB_\varepsilon(r) - \left(\int_0^1 \operatorname{sgn}\{B_\varepsilon(r)\}\, dr\right) B_\varepsilon(1)\right], \tag{64}$$

and $Z \sim N(0,1)$ is independent of $\mathcal{K}_4$.

Case 5: C5. We have $\rho_T = 1 + c/T^a$ with $c > 0$ and $0 < a < 1$. Define $\mathcal{J}_c^+(r) := \int_0^r e^{c(r-s)}\, dB_\varepsilon(s)$ and $S_c := \mathrm{sgn}(\int_0^\infty e^{-cs}\, dB_\varepsilon(s))$. As derived in the Proof of Theorem 5, we have

$$\begin{aligned} &\frac{1}{T^a}\sum_{t=1}^{T} |w_t - \mathrm{sgn}(X_T)| = O_p(1), \qquad \overline{w}_T - \mathrm{sgn}(X_T) = O_p(T^{a-1}), \\ &\frac{1}{T^a}\sum_{t=1}^{T} \{w_t - \mathrm{sgn}(X_T)\}^2 \Rightarrow \mathcal{V}_c^+, \qquad \frac{J_T}{T^a} \Rightarrow \mathcal{V}_c^+, \end{aligned} \tag{65}$$

where $\mathcal{V}_c^+ := \int_0^\infty \{\mathrm{sgn}(\mathcal{J}_c^+(r)) - S_c\}^2\, dr \in (0,\infty)$, a.s.

Applying (42) on the T^a time scale gives, we have on every compact interval that

$$\left(\frac{X_{\lfloor T^a r \rfloor}}{\sigma_\varepsilon T^{a/2}}, \frac{1}{\sigma T^{a/2}} \sum_{t=1}^{\lfloor T^a r \rfloor} U_t \right) \Longrightarrow \big(\mathcal{J}_c^+(r), B_U(r)\big). \tag{66}$$

The compact-interval smoothing argument used for C3 applies here as well. To pass to the whole sample, put $m = \lfloor T^a R \rfloor$. As in the proof of Theorem 5, we localize to $\{|\rho_T^{-m} X_m| > 2\epsilon T^{a/2}\}$ and stop when the discounted increment from time m first exceeds $\epsilon T^{a/2}$ in absolute value. The probabilities excluded by this localization can be made arbitrarily small by taking $T \to \infty$, then $R \to \infty$, and finally $\epsilon \downarrow 0$.

Before stopping, freeze the sign at the $\mathcal{G}_m$-measurable $\mathrm{sgn}(X_m)$. The predictable stopped weights satisfy

$$|w_t - \mathrm{sgn}(X_m)| \le C_\epsilon T^{-a\gamma/2} \rho_T^{-\gamma(t-1)}, \qquad t > m. \tag{67}$$

By Itô isometry, the second moment of the stopped score tail, after division by $T^{a/2}$, is bounded by

$$C_\epsilon T^{-a-a\gamma} \sum_{t=m+1}^{T} \rho_T^{-2\gamma(t-1)} \mathbb{E}(\sigma_t^2) \le C_\epsilon T^{-a\gamma} \rho_T^{-2\gamma m} = o(1). \tag{68}$$

Note that this argument uses only $\mathbb{E}(\sigma_t^2) = \sigma^2$; it does not require conditional homoscedasticity.

On the corresponding no-exit event, $\mathrm{sgn}(X_T) = \mathrm{sgn}(X_m)$. For $j = 2$ and $j = 2+q$, (67) also gives

$$\frac{1}{T^a}\sum_{t=m+1}^{T} |U_t|^j |w_t - \mathrm{sgn}(X_T)|^j \le C_\epsilon T^{-a-aj\gamma/2} \sum_{t=m+1}^{T} |U_t|^j \rho_T^{-j\gamma(t-1)}. \tag{69}$$

The expectation of the right-hand side is at most $C_\epsilon T^{-aj\gamma/2} \rho_T^{-j\gamma m} = o(1)$. The same second-moment bound holds with U_t^2 replaced by σ_t^2.

Equations (66)-(68), together with terminal-sign centering, imply that

$$\frac{1}{\sigma T^{a/2}} \sum_{t=1}^{T} U_t \{w_t - \mathrm{sgn}(X_T)\} \Rightarrow \int_0^\infty \{\mathrm{sgn}(\mathcal{J}_c^+(r)) - S_c\}\, dB_U(r). \tag{70}$$

For each Brownian motion B used here, the centered integral means

$$\int_0^\infty \{\mathrm{sgn}(\mathcal{J}_c^+(r)) - S_c\}\, dB(r) := \lim_{R\to\infty}\left[\int_0^R \mathrm{sgn}(\mathcal{J}_c^+(r))\, dB(r) - S_c B(R)\right]. \tag{71}$$

The truncated differences eventually stabilize almost surely, because the discounted diffusion converges to a nonzero Gaussian limit.

For the denominator, apply (57) with T^a in place of T on each fixed interval $[0, R]$. Smoothing the weights and using zero occupation time at zero yield

$$\frac{1}{T^a}\sum_{t=1}^{m}(U_t^2 - \sigma^2)\{w_t - \mathrm{sgn}(X_T)\}^2 \xrightarrow{p} 0. \tag{72}$$

The same statement holds with U_t^2 replaced by σ_t^2. The tail bounds in (69) and (65) then give, over the full sample, $T^{-a}\sum_{t=1}^{T}(U_t^2 - \sigma^2)\{w_t - \mathrm{sgn}(X_T)\}^2 = o_p(1)$ and $T^{-a}\sum_{t=1}^{T}(\sigma_t^2 - \sigma^2)\{w_t - \mathrm{sgn}(X_T)\}^2 = o_p(1)$. Replacing terminal-sign centering by sample-mean centering changes (70) by $((\overline{w}_T - \mathrm{sgn}(X_T))/T^{a/2})\sum_{t=1}^{T} U_t = O_p(T^{(a-1)/2}) = o_p(1)$.

For the second moment, using the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\frac{1}{T^a}\left|\sum_{t=1}^{T} U_t^2\left[(w_t^c)^2 - \{w_t - \mathrm{sgn}(X_T)\}^2\right]\right| \\
&\le \frac{2|\overline{w}_T - \mathrm{sgn}(X_T)|}{T^a}\left[\sum_t U_t^2\{w_t - \mathrm{sgn}(X_T)\}^2 \sum_t U_t^2\right]^{1/2} + \frac{\{\overline{w}_T - \mathrm{sgn}(X_T)\}^2}{T^a}\sum_t U_t^2 \\
&= O_p(T^{(a-1)/2}) + O_p(T^{a-1}) = o_p(1).
\end{aligned} \tag{73}$$

The identical argument applies with U_t^2 replaced by σ_t^2. Hence $T^{1-a}(V_T - \eta_T^2) = o_p(1)$, and $T^{1-a}\eta_T^2 - \sigma^2\frac{J_T}{T^a} = o_p(1)$. In particular, $V_T/\eta_T^2 \xrightarrow{p} 1$, even though $\eta_T^2 \xrightarrow{p} 0$.

Combining these results with (43) yields

$$\frac{A_T}{\sqrt{V_T}} \Rightarrow \sqrt{1-\lambda_U^2}\, Z + \lambda_U \mathcal{K}_5, \tag{74}$$

where

$$\mathcal{K}_5 := (\mathcal{V}_c^+)^{-1/2}\int_0^\infty \{\mathrm{sgn}(\mathcal{J}_c^+(r)) - S_c\}\, dB_\varepsilon(r). \tag{75}$$

Conditional on the entire path of B_ε, the corresponding $B_\perp$ integral divided by $\sqrt{\mathcal{V}_c^+}$ is $N(0,1)$. Thus, $Z \sim N(0,1)$ is independent of $\mathcal{K}_5$.

We now verify the maximum-score condition and complete the empirical likelihood argument. In C1-C4, bounded weights and the moment assumption imply $\sum_t |U_t w_t^c|^{2+q} = O_p(T)$. In C5, the first $\lfloor T^a R\rfloor$ terms have total $2+q$ moment $O(T^a)$, and (69) controls the remaining terms after terminal-sign centering. Moreover,

$$|\overline{w}_T - \mathrm{sgn}(X_T)|^{2+q}\sum_{t=1}^{T}|U_t|^{2+q} = O_p\Big(T^{(a-1)(2+q)+1}\Big) = o_p(T^a). \tag{76}$$

It follows that $\sum_t |U_t w_t^c|^{2+q} = O_p(T^a)$ in C5. Together with the regime-specific limits of J_T, this gives

$$\frac{\max_{1\le t\le T} |U_t w_t^c|}{\sqrt{J_T}} = o_p(1), \qquad \frac{\max_{1\le t\le T} (w_t^c)^2}{J_T} = o_p(1). \tag{77}$$

The preceding variance limits also imply $TV_T/J_T = O_p(1)$ and $J_T/(TV_T) = O_p(1)$.

Only consistency of the first-stage intercept estimator is needed for the remaining plug-in argument. Using $\widetilde{\alpha} - \alpha = o_p(1)$, we have

$$\frac{1}{J_T}\left|\sum_t \widetilde{Z}_t(\beta_0)^2 - \sum_t U_t^2 (w_t^c)^2\right| \le 2|\widetilde{\alpha} - \alpha| \left[\frac{1}{J_T}\sum_t U_t^2 (w_t^c)^2\right]^{1/2} + (\widetilde{\alpha} - \alpha)^2 = o_p(1). \tag{78}$$

Also, by (77),

$$\frac{\max_t |\widetilde{Z}_t(\beta_0)|}{\sqrt{J_T}} \le \frac{\max_t |U_t w_t^c|}{\sqrt{J_T}} + |\widetilde{\alpha} - \alpha| \frac{\max_t |w_t^c|}{\sqrt{J_T}} = o_p(1). \tag{79}$$

Since the numerator in (34) is unchanged exactly,

$$\frac{\sum_t \widetilde{Z}_t(\beta_0)}{\{\sum_t \widetilde{Z}_t(\beta_0)^2\}^{1/2}} = \frac{A_T}{\sqrt{V_T}} + o_p(1) = O_p(1). \tag{80}$$

Equations (79)–(80) imply that zero lies in the interior of the scalar cnvex hull of the estimating functions with probability tending to one.

Indeed, if all nonzero estimating functions had the same sign, their standardized sum in absolute value would be at least $\{\sum_t \widetilde{Z}_t(\beta_0)^2\}^{1/2} / \max_t |\widetilde{Z}_t(\beta_0)|$, which diverges in probability. This contradicts (80). The multiplier equation therefore yields $\widetilde{\lambda} = \frac{\sum_t \widetilde{Z}_t(\beta_0)}{\sum_t \widetilde{Z}_t(\beta_0)^2} + o_p(J_T^{-1/2})$, and $\max_t |\widetilde{\lambda}\widetilde{Z}_t(\beta_0)| = o_p(1)$. The Taylor remainder is bounded by

$$C\sum_{t=1}^{T} |\widetilde{\lambda}\widetilde{Z}_t(\beta_0)|^3 \le C \max_{1\le t\le T} |\widetilde{\lambda}\widetilde{Z}_t(\beta_0)| \widetilde{\lambda}^2 \sum_{t=1}^{T} \widetilde{Z}_t(\beta_0)^2 = o_p(1).$$

Consequently,

$$\widetilde{\ell}_T(\beta_0) = 2\sum_{t=1}^{T} \log\{1 + \widetilde{\lambda}\widetilde{Z}_t(\beta_0)\} \frac{\{\sum_t \widetilde{Z}_t(\beta_0)\}^2}{\sum_t \widetilde{Z}_t(\beta_0)^2} + o_p(1) = \frac{A_T^2}{V_T} + o_p(1). \tag{81}$$

Thus, for general λ_U,

$$\widetilde{\ell}_T(\beta_0) \Rightarrow \begin{cases} Z^2, & \text{C1 and C2}, \\ \left(\sqrt{1-\lambda_U^2}\, Z + \lambda_U \mathcal{K}_j\right)^2, & \text{C}j, \quad j = 3,4,5, \end{cases} \tag{82}$$

where, in each case, $Z \sim N(0,1)$ is independent of the corresponding $\mathcal{K}_j$. This completes the proof. Under the condition $\mathbb{E}[U_t \varepsilon_t] = 0$, we have $\lambda_U = 0$ and $\lambda_U \mathcal{K}_j$ vanishes. Therefore, $\widetilde{\ell}_T(\beta_0) \Rightarrow \chi_1^2$ in all five persistence classes, which proves Corollary 1. □

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 $\psi_\tau(u) = \tau - 1(u < 0)$, $w_t := w(X_{t-1})$, $\overline{w}_T := T^{-1}\sum_{s=1}^T w_s$, $w_t^c := w_t - \overline{w}_T$, and $J_T := \sum_{t=1}^T (w_t^c)^2$. Throughout, $\theta = 0$. Under the null hypothesis $H_0 : \beta_\tau = \beta_{\tau,0}$, given the estimator (17) and the two-stage intercept-adjusted response $\widetilde{Y}_t := Y_t - \widetilde{\alpha}_\tau$ we have

$$\widetilde{\xi}_{t,\tau}(\beta_{\tau,0}) \;:=\; \psi_\tau(\widetilde{Y}_t - \beta_{\tau,0}X_{t-1})w_t^c \;=\; \psi_\tau(Y_t - \widetilde{\alpha}_\tau - \beta_{\tau,0}X_{t-1})w_t^c. \tag{83}$$

Under H_0, we have $Y_t - \widetilde{\alpha}_\tau - \beta_{\tau,0}X_{t-1} = U_{t,\tau} - (\widetilde{\alpha}_\tau - \alpha_\tau)$, and hence

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^T \widetilde{\xi}_{t,\tau}(\beta_{\tau,0}) &= \frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_\tau(U_{t,\tau})w_t^c \;+\; \frac{1}{\sqrt{T}}\sum_{t=1}^T \Big\{\psi_\tau\big(U_{t,\tau} - (\widetilde{\alpha}_\tau - \alpha_\tau)\big) - \psi_\tau(U_{t,\tau})\Big\}w_t^c \\ &=:\; A_T + R_T. \end{aligned} \tag{84}$$

We first establish a Bahadur representation. Since $\widetilde{\alpha}_\tau$ is the sample τ-quantile of $\{\alpha_\tau + U_{t,\tau}\}_{t=1}^T$ under H_0, we have $T^{-1/2}\sum_t \psi_\tau(U_{t,\tau} - (\widetilde{\alpha}_\tau - \alpha_\tau)) = o_p(1)$. For a fixed $d \in \mathbb{R}$, define $D_t(d) := \psi_\tau(U_{t,\tau} - d) - \psi_\tau(U_{t,\tau})$ and $m_{t,\tau}(d) := \mathbb{E}\big[D_t(d)|\mathcal{F}_{t-1}\big]$. Then, $D_t(d) - m_{t,\tau}(d)$ is a bounded martingale difference, and the local boundedness of the conditional density gives, uniformly for sufficiently small d, $\mathbb{E}[\{D_t(d) - m_{t,\tau}(d)\}^2|\mathcal{F}_{t-1}] \le C|d|$.

By Freedman's martingale inequality, see e.g. Tropp (2011, Theorem 1.1), applied on a $T^{-3/4}$-grid together with the monotonicity of the indicator increment, it follows that

$$\sup_{|d|\le K/\sqrt{T}} \left| \frac{1}{\sqrt{T}}\sum_{t=1}^T \{D_t(d) - m_{t,\tau}(d)\} \right| = o_p(1) \tag{85}$$

Meanwhile, Assumption 3 implies $m_{t,\tau}(d) = -f_\tau(0)d + O(d^2)$ uniformly in t, and hence, uniformly for $|d| \le K/\sqrt{T}$ for any $K < \infty$, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_\tau(U_{t,\tau} - d) \;=\; \frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_\tau(U_{t,\tau}) - f_\tau(0)\sqrt{T}d + o_p(1). \tag{86}$$

Since $T^{-1/2}\sum_t \psi_\tau(U_{t,\tau})$ is bounded in probability, evaluating (86) at $d = \pm K/\sqrt{T}$ and using monotonicity first gives $\widetilde{\alpha}_\tau - \alpha_\tau = O_p(T^{-1/2})$. Therefore, substituting $d = \widetilde{\alpha}_\tau - \alpha_\tau$ into (86) yields

$$\sqrt{T}\,(\widetilde{\alpha}_\tau - \alpha_\tau) \;=\; \frac{1}{f_\tau(0)}\frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_\tau(U_{t,\tau}) + o_p(1). \tag{87}$$

Now write

$$R_T = \frac{1}{\sqrt{T}}\sum_{t=1}^T w_t^c\Big[D_t(\widetilde{\alpha}_\tau - \alpha_\tau) - m_{t,\tau}(\widetilde{\alpha}_\tau - \alpha_\tau)\Big] \;+\; \frac{1}{\sqrt{T}}\sum_{t=1}^T m_{t,\tau}(\widetilde{\alpha}_\tau - \alpha_\tau)w_t^c = R_{T,1} + R_{T,2}.$$

By Assumption 3, there exists a constant $f_\tau(0) \in (0, \infty)$ and $\epsilon > 0$ such that, uniformly in t,

$F_{t,\tau}(d|\mathcal{F}_{t-1}) = \tau + f_\tau(0)d + O(d^2)$ a.s. for $|d| \le \epsilon$, so we have $m_{t,\tau}(d) = -f_\tau(0)d + r_{t,\tau}(d)$ and $|r_{t,\tau}(d)| \le Cd^2$, uniformly in t and locally in d.

Since $\sum_{t=1}^T w_t^c = 0$ it follows that

$$\begin{aligned} J_T^{-1/2}\left|\sum_{t=1}^T m_{t,\tau}(\widetilde{\alpha}_\tau - \alpha_\tau)w_t^c\right| &= J_T^{-1/2}\left|-f_\tau(0)(\widetilde{\alpha}_\tau - \alpha_\tau)\sum_{t=1}^T w_t^c + \sum_{t=1}^T r_{t,\tau}(\widetilde{\alpha}_\tau - \alpha_\tau)w_t^c\right| \\ &= J_T^{-1/2}\left|\sum_{t=1}^T r_{t,\tau}(\widetilde{\alpha}_\tau - \alpha_\tau)w_t^c\right| \le C(\widetilde{\alpha}_\tau - \alpha_\tau)^2\frac{\sum_{t=1}^T |w_t^c|}{\sqrt{J_T}} \\ &\le C\sqrt{T}(\widetilde{\alpha}_\tau - \alpha_\tau)^2 = O_p(T^{-1/2}) = o_p(1), \end{aligned} \tag{88}$$

by the Cauchy-Schwarz inequality. Note that as we show below, J_T is of order T for cases C1-C4 and of order T^a (with $a < 1$) for case C5.

We now prove the Lee (2016) analogue of stochastic equicontinuity under persistence-specific normalization. Lemma A.3 of Lee (2016) cannot be directly applied because $\overline{w}_T$ is not $\mathcal{F}_{t-1}$-measurable. For deterministic $\nu \in [-\|w\|_\infty, \|w\|_\infty]$, write $J_T(\nu) := \sum_{t=1}^T (w_t - \nu)^2$. Since $J_T(\nu) = J_T + T(\nu - \overline{w}_T)^2 \ge J_T$ and $J_T(\overline{w}_T) = J_T$, it suffices to obtain bounds uniformly in ν. For each fixed ν and d, $\sum_t (w_t - \nu)\{D_t(d) - m_{t,\tau}(d)\}$ is a bounded martingale sum whose predictable quadratic variation is bounded by $C|d|J_T(\nu)$.

The regime-specific weight limits established below imply $J_T/(\log T)^2 \to^p \infty$. To handle the random normalizer, we apply Freedman's inequality on dyadic bands of $J_T(\nu) \ge J_T$, followed by a union bound over deterministic grids in (ν, d) with mesh T^{-3}. The bounds extend between grid points using the local conditional density bound and monotonicity of the indicator increments. Thus,

$$\sup_{\nu\in[-\|w\|_\infty,\|w\|_\infty]}\sup_{|d|\le K/\sqrt{T}}\frac{\left|\sum_{t=1}^T (w_t - \nu)\{D_t(d) - m_{t,\tau}(d)\}\right|}{\sqrt{J_T(\nu)}} = o_p(1). \tag{89}$$

The same argument applied to $(w_t - \nu)^2[1\{|U_{t,\tau}| \le x\} - \mathbb{P}(|U_{t,\tau} \le x|\mathcal{F}_{t-1})]$, together with monotonicity in x yields

$$\sup_{\nu\in[-\|w\|_\infty,\|w\|_\infty]}\sup_{|d|\le K/\sqrt{T}} J_T(\nu)^{-1}\sum_{t=1}^T (w_t - \nu)^2 1\{|U_{t,\tau} \le |d|\} = o_p(1). \tag{90}$$

Evaluating the two statements (89) and (90) at $\nu = \overline{w}_T$ gives $\sup_{|d|\le K/\sqrt{T}} J_T^{-1}\sum_{t=1}^T (w_t^c)^2 \cdot 1\{|U_{t,\tau}| \le |d|\} = o_p(1)$ and

$$\left|\sum_{t=1}^T w_t^c\Big[D_t(\widetilde{\alpha}_\tau - \alpha_\tau) - m_{t,\tau}(\widetilde{\alpha}_\tau - \alpha_\tau)\Big]\right| = o_p(\sqrt{J_T}). \tag{91}$$

Combining (88) and (91), we obtain $\sum_{t=1}^T \widetilde{\xi}_{t,\tau}(\beta_{\tau,0}) = \sum_{t=1}^T \psi_\tau(U_{t,\tau})w_t^c + o_p(\sqrt{J_T})$. Similarly,

since $|\psi_\tau(U_{t,\tau} - (\widetilde{\alpha}_\tau - \alpha_\tau))^2 - \psi_\tau(U_{t,\tau})^2| \leq |2\tau - 1| \cdot 1\{|U_{t,\tau}| \leq |\widetilde{\alpha}_\tau - \alpha_\tau|\}$, it follows that

$$\sum_{t=1}^{T} \widetilde{\xi}_{t,\tau}(\beta_{\tau,0})^2 = \sum_{t=1}^{T} \psi_\tau(U_{t,\tau})^2 (w_t^c)^2 + o_p(J_T). \tag{92}$$

Since $(\psi_\tau(U_{t,\tau}), \varepsilon_t)'$ is a stationary martingale difference, by the same functional central limit theorem argument as in equation (2.7) of Lee (2016) we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \begin{pmatrix} \psi_\tau(U_{t,\tau}) \\ \varepsilon_t \end{pmatrix} \Rightarrow \begin{pmatrix} \sqrt{\tau(1-\tau)}\, B_{\psi,\tau}(r) \\ \sigma_\varepsilon B_\varepsilon(r) \end{pmatrix}, \tag{93}$$

where $B_{\psi,\tau}$ and B_ε are standard Brownian motions satisfying $\mathbb{E}\big[B_{\psi,\tau}(r) B_\varepsilon(r)\big] = \lambda_\tau r$ and

$$\lambda_\tau := \frac{\mathbb{E}[\psi_\tau(U_{t,\tau})\varepsilon_t]}{(\sigma_\varepsilon \sqrt{\tau(1-\tau)})} = \frac{\Sigma_{\psi_\tau \varepsilon}}{(\sigma_\varepsilon \sqrt{\tau(1-\tau)}}. \tag{94}$$

Consequently, on an enlarged probability space, we may write $B_{\psi,\tau}(r) = \sqrt{1-\lambda_\tau^2}\, B_\perp(r) + \lambda_\tau B_\varepsilon(r)$, where $B_\perp$ is a standard Brownian motion independent of B_ε. It is worth noting that when $\mathbb{E}[\psi_\tau(U_{t,\tau})\varepsilon_t] = 0$, we have $\lambda_\tau = 0$ and $B_{\psi,\tau}$ and B_ε are independent.

We now consider the self-normalizer. Note that $\psi_\tau(U_{t,\tau})^2 = \tau(1-\tau) + (2\tau - 1)\psi_\tau(U_{t,\tau})$, and that $\sum_t \psi_\tau(U_{t,\tau})(w_t^c)^2 = \sum_t \psi_\tau(U_{t,\tau}) w_t^2 - 2\overline{w}_T \sum_t \psi_\tau(U_{t,\tau}) w_t + \overline{w}_T^2 \sum_t \psi_\tau(U_{t,\tau})$. All three terms in the summation are $O_p(\sqrt{T})$ in C1-C4, since $1, w_t, w_t^2$ are bounded and $\mathcal{F}_{t-1}$-measurable. As for C5, for deterministic ν, we let $Q_T(\nu) := \sum_{t=1}^T \psi_\tau(U_{t,\tau})(w_t - \nu)^2$. This is a martingale sum with bounded increments and predictive quadratic variation

$$\langle Q(\nu) \rangle_T = \tau(1-\tau) \sum_{t=1}^{T} (w_t - \nu)^4 \leq C J_T(\nu).$$

The same localization, dyadic decomposition, and grid argument therefore yield

$$\sup_{\nu \in [-\|w\|_\infty, \|w\|_\infty]} \frac{|Q_T(\nu)|}{J_T(\nu)} = o_p(1).$$

Evaluating at $\nu = \overline{w}_T$ yields $\sum_t \psi_t(U_{t,\tau})(w_t^c)^2 = o_p(J_T)$. Consequently, we have

$$\begin{aligned} \sum_{t=1}^{T} \psi_\tau(U_{t,\tau})(w_t^c)^2 &= o_p(J_T) \\ \sum_{t=1}^{T} \psi_\tau(U_{t,\tau})^2 (w_t^c)^2 &= \tau(1-\tau) J_T + o_p(J_T). \end{aligned}$$

Together with (92), this yields $\sum_{t=1}^{T}\widetilde{\xi}_{t,\tau}(\beta_{\tau,0})^2 = \tau(1-\tau)J_T + o_p(J_T)$, and hence

$$\frac{\sum_{t=1}^{T}\widetilde{\xi}_{t,\tau}(\beta_{\tau,0})}{\left[\sum_{t=1}^{T}\widetilde{\xi}_{t,\tau}(\beta_{\tau,0})^2\right]^{1/2}} = \frac{\sum_{t=1}^{T}\psi_\tau(U_{t,\tau})w_t^c}{\{\tau(1-\tau)J_T\}^{1/2}} + o_p(1) =: Z_{T,\tau} + o_p(1). \tag{95}$$

It remains to establish the limiting distribution of $Z_{T,\tau}$ in each persistence class, since the boundedness of ψ_τ and w, $\max_{1\le t\le T}(w_t^c)^2/J_T = o_p(1)$, and Taylor expansion gives

$$\widetilde{\ell}_{T,\tau}(\beta_{\tau,0}) = 2\sum_{t=1}^{T}\log\{1+\lambda_\tau\widetilde{\xi}_{t,\tau}(\beta_{\tau,0})\} = Z_{T,\tau}^2 + o_p(1). \tag{96}$$

In C1, let $X_t' := \sum_{j=0}^{\infty}\rho^j\varepsilon_{t-j}$ denote the stationary solution of (2). Given that $\mathbb{E}|\varepsilon_t| < \infty$, the series $\mathbb{E}[\sum_{j=0}^{\infty}|\rho|^j|\varepsilon_{t-j}|] < \infty$. Since $X_t - X_t' = \rho^t(X_0 - X_0')$ with $|\rho| < 1$, and both X_0 and X_0' are bounded in probability, the initial transient vanishes geometrically. By the boundedness and uniform continuity of w and the Cesàro Theorem, $T^{-1}\sum_t(w(X_{t-1}) - w(X_{t-1}'))^2 = o_p(1)$. Hence, due to stationarity and ergodicity of ε_t (and hence of X_t') and by the Cauchy-Schwartz inequality, $\overline{w}_T \to^p \mathbb{E}[w(X_0')]$ and

$$\frac{J_T}{T} = \frac{1}{T}\sum_{t=1}^{T}(w_t^c)^2 \xrightarrow{p} \mathbb{E}[\{w(X_0') - \mathbb{E}[w(X_0')]\}^2] > 0,$$

where the strict positivity follows from the maintained condition. Since $w_t - \mathbb{E}[w(X_0')]$ is bounded and $\mathcal{F}_{t-1}$-measurable, and the Lindeberg condition immediately follows. Lastly, given that $T^{-1}\sum_t \mathbb{E}[\psi_\tau(U_{t,\tau})^2(w_t - \mathbb{E}[w(X_0')])^2|\mathcal{F}_{t-1}] = \tau(1-\tau)\{J_T/T + (\overline{w}_T - \mathbb{E}[w(X_0')])^2\}$, it follows that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\tau(U_{t,\tau})\{w_t - \mathbb{E}[w(X_0')]\} \Rightarrow N\big(0, \tau(1-\tau)\mathbb{E}[\{w(X_0') - \mathbb{E}[w(X_0')]\}^2]\big)$$

by the martingale central limit theorem, e.g. Hall and Heyde (1980). The centering remainder $(\overline{w}_T - \mathbb{E}[w(X_0')])\sum_{t=1}^{T}\psi_\tau(U_{t,\tau}) = o_p(1)O_p(\sqrt{T}) = o_p(\sqrt{T})$. Therefore, $Z_{T,\tau}$ converges in distribution to standard normal and $\widetilde{\ell}_{T,\tau}(\beta_{\tau,0}) \Rightarrow \chi_1^2$ by (96), as required.

In C2, $\rho = \rho_T = 1 + cT^{-a}$ with $c < 0$ and $0 < a < 1$. In light of the initial condition, we define the stationary companion $X_{T,t}' := \sum_{j=0}^{\infty}\rho^j\varepsilon_{t-j}$, and note that $s_T^2 = \mathbb{E}[\varepsilon_0^2]/(1-\rho_T^2) = O(T^a)$. Since for $b_{T,j} := (1-\rho_T^2)^{1/2}\rho_T^j 1\{j \ge 0\}$, $j \in \mathbb{Z}$ we have $\sum_{j\in\mathbb{Z}} b_{T,j}^2 = 1$ and $\sum_j(b_{T,j} - b_{T,j-1})^2 = 2(1-\rho_T) \to 0$, Lemma 2.1(ii) and Theorem 2 of Abadir, Distaso, Giraitis, and Koul (2014) yield $X_{T,0}'/s_T \Rightarrow N(0,1)$.

For each fixed $r \in (0,1)$, let $k_T = \lfloor Tr \rfloor$. The AR recursion, martingale difference properties, and the fact that $\log(1-x) \le -x$ yield $\mathrm{Cov}(X_{T,0}'/s_T, X_{T,k_T}'/s_T) = \rho_T^{k_T} \le \exp(ck_T/T^a) \to 0$. Since every fixed nonzero linear combination of these two normalized sums satisfies the same coefficient-smoothness condition after variance normalization, the Cramér-Wold device yields

$$(X_{T,0}'/s_T, X_{T,k_T}'/s_T) \Rightarrow (Z_1, Z_2)$$

where Z_1 and Z_2 are independent standard normal variables.

Let $q_T(k) := \mathbb{E}[w(X'_{T,0})w(X'_{T,k})]$. Since $s_T \to \infty$, boundedness and saturation of w, together with the preceding Gaussian limits, imply that $\mathbb{E}|(w(X'_{T,0})^2 - 1| \to 0$. Also, using stationarity, it follows that $|\mathbb{E}[w(X'_{T,0})w(X'_{T,k}) - \text{sgn}(X'_{T,0})\text{sgn}(X'_{T,k_T})]| \leq C\mathbb{E}|w(X'_{T,0}) - \text{sgn}(X'_{T,0})|$. Therefore, noting that $h(x,y) := \text{sgn}(x)\text{sgn}(y)$ is bounded, we see that $\mathbb{E}[\text{sgn}(X'_{T,0})\text{sgn}(X'_{T,k_T})] \to 0$, yielding $q_T(\lfloor Tr \rfloor) \to 0$ for every $r \in (0,1)$.

Since by stationarity we have $\mathbb{E}|T^{-1}\sum_t (w'_{T,t-1})^2 - 1| \leq \mathbb{E}|(w'_{T,0})^2 - 1| \to 0$, bounded convergence gives

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T} w'_{T,t-1}\right)^2\right] &= \frac{q_T(0)}{T} + \frac{2}{T}\sum_{k=1}^{T-1}\left(1 - \frac{k}{T}\right)q_T(k) \\ &= 2\int_0^1 \left(1 - \frac{\lfloor Tr \rfloor}{T}\right) q_T(\lfloor Tr \rfloor)dr - \frac{q_T(0)}{T} \to 0.\end{aligned}$$

Now, choose $m_T = \lceil KT^a \log T \rceil$, $K > a/(-2c)$. Then, since $m_T/T \to 0$, $X'_{T,0} = O_p(T^{a/2})$, and $\sup_{t \geq m_T} |X_t - X'_{T,t} \leq \rho_T^{m_T}|X_0 - X'_{T,0}| = o_p(1)$, boundedness and uniform continuity of w imply that for $g = w$, and w^2, we have $T^{-1}\sum_{t=1}^T |g(X_{t-1}) - g(X'_{t-1})| = o_p(1)$. Hence,

$$\overline{w}_T = o_p(1), \quad \frac{1}{T}\sum_{t=1}^{T} w_t^2 \xrightarrow{p} 1, \quad \frac{J_T}{T} = \frac{1}{T}\sum_{t=1}^{T} w_t^2 - \overline{w}_T^2 \xrightarrow{p} 1.$$

Finally, $\psi_\tau(U_{t,\tau})w_t$ is a martingale difference, and

$$\frac{1}{\tau(1-\tau)T}\sum_{t=1}^{T} \mathbb{E}[\psi_\tau(U_{t,\tau})^2 w_t^2 | \mathcal{F}_{t-1}] = \frac{1}{T}\sum_{t=1}^{T} w_t^2 \xrightarrow{p} 1.$$

Because w_t is bounded and $\mathcal{F}_{t-1}$-measurable, the conditional Lindeberg condition follows, and the martingale central limit theorem yields $\frac{1}{\sqrt{\tau(1-\tau)T}}\sum_{t=1}^T \psi_\tau(U_{t,\tau})w_t \Rightarrow N(0,1)$. Finally, since $\overline{w}_T \frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_\tau(U_{t,\tau}) = o_p(1)O_p(1) = o_p(1)$, we have

$$Z_{T,\tau} = (J_T/T)^{-1/2}\left[\frac{1}{\sqrt{\tau(1-\tau)T}}\sum_{t=1}^{T}\psi_\tau(U_{t,\tau})w_t + o_p(1)\right] \Rightarrow N(0,1), \tag{97}$$

and hence, in view of (96), $\widetilde{\ell}_{T,\tau}(\beta_{\tau,0}) = Z_{T,\tau}^2 + o_p(1) \Rightarrow \chi_1^2$, as required.

In C3, $\rho_T = 1 + c/T$ with $c \neq 0$. By (93) and standard unit root limit theory, we have $X_{\lfloor Tr \rfloor}/(\sigma_\varepsilon\sqrt{T}) \Rightarrow \mathcal{J}_c(r) := \int_0^r e^{c(r-s)}\, dB_\varepsilon(s)$. Write $\mathcal{V}_c := 1 - [\int_0^1 \text{sgn}\{\mathcal{J}_c(r)\}\, dr]^2$.

Since $e^{-cr}\mathcal{J}_c(r)$ is a deterministically time-changed Brownian motion, $\mathcal{J}_c$ has zero occupation time at zero and spends positive time on both half-lines a.s. Hence, $|\int_0^1 \text{sgn}(\mathcal{J}_c(r))dr| < 1$ and $\mathcal{V}_c > 0$, a.s, including when $c = 0$. Throughout C3 and C4, centered-integrand expressions are understood as the following shorthand:

$$\int_0^1 \left[\text{sgn}\{\mathcal{J}_c(r)\} - \int_0^1 \text{sgn}\{\mathcal{J}_c(r)\}dr\right] dB(r) := \int_0^1 \text{sgn}\{\mathcal{J}_c(r)\}dB(r) - \int_0^1 \text{sgn}\{\mathcal{J}_c(r)\}drB(1),$$

for $B = B_{\psi,\tau}, B_\perp(r), B_\varepsilon$. The integral on the right is an ordinary Itô integral; the notion on the left does not assert predictability of the centered integrand.

Continuous mapping and the saturation property of w imply $\overline{w}_T \Rightarrow \int_0^1 \mathrm{sgn}\{\mathcal{J}_c(r)\}\,dr$, and $J_T/T \Rightarrow \mathcal{V}_c$. Let $g_\delta : \mathbb{R} \to [-1.1]$ be Lipschitz and equal to sgn outside $[-\delta, \delta]$. For each fixed $\delta > 0$ the convergence of stochastic integral applies to the smoothed weight, jointly with (93) and the normalized predictor: the score martingales $M_T(r) := (\tau(1-\tau)T)^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \psi_\tau(U_{t,\tau})$ have $\langle M_T \rangle_1 = 1$. By Itô isometry, saturation of w, and the zero occupation time of $\mathcal{J}_c$ at zero implies that

$$\frac{\sum_{t=1}^T \psi_\tau(U_{t,\tau}) w_t^c}{\sqrt{\tau(1-\tau)T}} \Rightarrow \int_0^1 \left\{ \mathrm{sgn}\{\mathcal{J}_c(r)\} - \int_0^1 \mathrm{sgn}\{\mathcal{J}_c(r)\}\,dr \right\} dB_{\psi,\tau}(r). \tag{98}$$

Therefore,

$$\begin{aligned} Z_{T,\tau} \Rightarrow\ & \sqrt{1-\lambda_\tau^2}\,\mathcal{V}_c^{-1/2} \int_0^1 \left\{ \mathrm{sgn}\{\mathcal{J}_c(r)\} - \int_0^1 \mathrm{sgn}\{\mathcal{J}_c(r)\}\,dr \right\} dB_\perp(r) \\ &+\ \lambda_\tau \mathcal{V}_c^{-1/2} \int_0^1 \left\{ \mathrm{sgn}\{\mathcal{J}_c(r)\} - \int_0^1 \mathrm{sgn}\{\mathcal{J}_c(r)\}\,dr \right\} dB_\varepsilon(r). \end{aligned} \tag{99}$$

where, conditional on B_ε, the first stochastic integral is $N(0,1)$ and its conditional distribution does not depend on B_ε. Define $\mathcal{K}_3 := \mathcal{V}_c^{-1/2}[\int_0^1 \mathrm{sgn}\{\mathcal{J}_c(r)\} dB_\varepsilon(r) - (\int_0^1 \mathrm{sgn}(\mathcal{J}_c(r))dr) B_\varepsilon(1)]$ Then,

$$Z_{T,\tau} \Rightarrow \sqrt{1-\lambda_\tau^2}\,Z + \lambda_\tau \mathcal{K}_3, \tag{100}$$

where $Z \sim N(0,1)$ is independent of $\mathcal{K}_3$. Therefore,

$$\widetilde{\ell}_{T,\tau}(\beta_{\tau,0}) \Rightarrow \left(\sqrt{1-\lambda_\tau^2}\,Z + \lambda_\tau \mathcal{K}_3 \right)^2 = Z^2 \tag{101}$$

when $\lambda_\tau = 0$, as required.

C4 is obtained from C3 by setting $c = 0$. In this case, $\mathcal{J}_0(r) = B_\varepsilon(r)$, and $J_T/T \Rightarrow 1 - [\int_0^1 \mathrm{sgn}\{B_\varepsilon(r)\} dr]^2$. The same argument as in C3 gives

$$Z_{T,\tau} \Rightarrow \sqrt{1-\lambda_\tau^2}\,Z + \lambda_\tau \mathcal{K}_4 \tag{102}$$

where $Z \sim N(0,1)$ is independent of $\mathcal{K}_4 := \mathcal{V}_0^{-1/2} \int_0^1 \{\mathrm{sgn}\{B_\varepsilon(r)\} - \int_0^1 \mathrm{sgn}\{B_\varepsilon(r)\} dr\}\, dB_\varepsilon(r)$. Therefore,

$$\widetilde{\ell}_{T,\tau}(\beta_{\tau,0}) \Rightarrow \left(\sqrt{1-\lambda_\tau^2}\,Z + \lambda_\tau \mathcal{K}_4 \right)^2 = Z^2 \tag{103}$$

when $\lambda_\tau = 0$, as required.

We lastly consider C5: $\rho_T = 1 + c/T^a$, where $c > 0$ and $0 < a < 1$. Define $\mathcal{J}_c^+(r) := \int_0^r e^{c(r-s)}\, dB_\varepsilon(s)$, and $S_c := \mathrm{sgn}(\int_0^\infty e^{-cs}\, dB_\varepsilon(s))$. The joint martingale FCLT underlying (93), applied on the T^a time scale, together with the AR recursion yields

$$\left(\frac{X_{\lfloor T^a r \rfloor}}{\sigma_\varepsilon T^{a/2}}, \frac{\sum_{t=1}^{\lfloor T^a r \rfloor} \psi_\tau(U_{t,\tau})}{\sqrt{\tau(1-\tau)T^a}} \right) \Rightarrow \left(\mathcal{J}_c^+(r), B_{\psi,\tau}(r) \right) \tag{104}$$

on every compact interval of $[0,\infty)$.

To control the tails, we put $m = \lfloor T^a R \rfloor$.The representation $\rho_T^{-t} X_t = X_0 + \sum_{j=1}^t \rho_T^{-j}\varepsilon_j$ and Doob's inequality give $\mathbb{E}[\sup_{m\le k\le T} |\{\rho_T^{-k}X_k - \rho_T^{-m}X_m\}(T^{-a/2})|^2] \le 4\sigma_\varepsilon^2 \rho_T^{-2m}\{T^a(\rho_T^2-1)\}^{-1}$, whose limit superior is bounded by Ce^{-2cR}. In particular, truncation of the discounted predictor gives $\mathrm{sgn}(X_T) \Rightarrow S_c$, jointly with the compact-interval limits.

For $\eta > 0$, localize to $\{|\rho_T^{-m}X_m| > 2\eta T^{a/2}\}$ and stop when the discounted increment from time m first exceeds $\eta T^{a/2}$ in absolute value. Before stopping, X_t has sign $\mathrm{sgn}(X_m)$ and $|X_t| \ge \eta T^{a/2}\rho_T^t$. Thus, on the corresponding no-exit event, the saturation assumption implies, for $j = 1, 2$,

$$\frac{1}{T^a}\sum_{t=m+1}^{T} |w_t - \mathrm{sgn}(X_T)|^j \le C_\eta T^{-ja\gamma/2}\rho_T^{-j\gamma m}.$$

For the score tail, freeze the sign at the $\mathcal{F}_m$-measurable $\mathrm{sgn}(X_m)$ and apply martingale isometry to the stopped sum. Its variance after $T^{-a/2}$ normalization is bounded by $C_\eta T^{-a\gamma}\rho_T^{-2\gamma m} = o(1)$. The excluded events have arbitrarily small probability by taking$T \to \infty$, then $R \to \infty$, and finally $\eta \downarrow 0$.

On each compact interval, the smoothing argument used in C3 applies to the uncentered weighted score. For the smoothed weights, stochastic-integral convergence follows from Kurtz and Protter (1991), since the normalized score martingales have uniformly bounded expected quadratic variation on each compact interval. The smoothing error vanishes by martingale isometry and the zero occupation time of $\mathcal{J}_c^+$ at zero. Terminal-sign centering is then handled algebraically by subtracting $\mathrm{sgn}(X_T)$ times the unweighted score.

Combining this argument with the preceding tail bounds gives the following joint limits:

$$\frac{1}{T^a}\sum_{t=1}^{T} |w_t - \mathrm{sgn}(X_T)| = O_p(1),$$

$$\frac{1}{T^a}\sum_{t=1}^{T} \{w_t - \mathrm{sgn}(X_T)\}^2 \Rightarrow \mathcal{V}_c^+ := \int_0^\infty \{\mathrm{sgn}(\mathcal{J}_c^+(r)) - S_c\}^2\, dr,$$

and

$$\frac{\sum_{t=1}^T \psi_\tau(U_{t,\tau})\{w_t - \mathrm{sgn}(X_T)\}}{\sqrt{\tau(1-\tau)T^a}} \Rightarrow \int_0^\infty \{\mathrm{sgn}(\mathcal{J}_c^+(r)) - S_c\}\, dB_{\psi,\tau}(r). \tag{105}$$

Here $\mathcal{V}_c^+ \in (0,\infty)$ almost surely: the discounted diffusion $e^{-cr}\mathcal{J}_c^+(r)$ oscillates on both sides of zero near the origin and converges almost surely to a nonzero Gaussian limit, so its sign eventually equals S_c. For each Brownian motion B appearing below, the centered integral is understood as $\int_0^\infty \{\mathrm{sgn}(\mathcal{J}_c^+(r)) - S_c\}\, dB(r) := \lim_{R\to\infty}[\int_0^R \mathrm{sgn}(\mathcal{J}_c^+(r))\, dB(r) - S_c B(R)]$.

These truncated differences eventually stabilize almost surely; this notation does not assert predictability of the terminal-sign-centered integrand. Since $\overline{w}_T - \mathrm{sgn}(X_T) = T^{-1}\sum_{t=1}^T\{w_t - \mathrm{sgn}(X_T)\} = O_p(T^{a-1})$, we have $\frac{J_T}{T^a} = \frac{1}{T^a}\sum_{t=1}^T\{w_t - \mathrm{sgn}(X_T)\}^2 - T^{1-a}\{\overline{w}_T - \mathrm{sgn}(X_T)\}^2 \Rightarrow \mathcal{V}_c^+$. Moreover, because $\sum_t \psi_\tau(U_{t,\tau}) = O_p(\sqrt{T})$, replacing $w_t - \mathrm{sgn}(X_T)$ by w_t^c in (105) changes its left-hand side by $O_p(T^{(a-1)/2}) = o_p(1)$.

Consequently, we have$Z_{T,\tau} \Rightarrow \sqrt{1-\lambda_\tau^2}\, Z + \lambda_\tau \mathcal{K}_5$, where $\mathcal{K}_5 := (\mathcal{V}_c^+)^{-1/2}\int_0^\infty\{\mathrm{sgn}(\mathcal{J}_c^+(r)) -$

$S_c\} \, dB_\varepsilon(r)$. Conditional on the entire path of B_ε, the $(\mathcal{V}_c^+)^{-1/2}$-normalized integral with respect to $B_\perp$ is $N(0,1)$. Thus $Z \sim N(0,1)$ is independent of $\mathcal{K}_5$. By (96),

$$\widetilde{\ell}_{T,\tau}(\beta_{\tau,0}) \Rightarrow \left(\sqrt{1-\lambda_\tau^2}\, Z + \lambda_\tau \mathcal{K}_5\right)^2. \tag{106}$$

Under the maintained condition $\lambda_\tau = 0$, this reduces to the required χ_1^2 limit. □

References

Abadir, K. M., W. Distaso, L. Giraitis, and H. L. Koul (2014). Asymptotic normality for weighted sums of linear processes. *Econometric Theory*, 30(1), 252-284.

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.

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

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

Kurtz, T. G. and P. Protter. (1991). Weak Limit Theorems for STochastic Integrals and Stochastic Differential Equations. *The Annals of Probability*, 19(3), 1035-1070.

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.

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

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

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

Tropp, J. A. (2011). Freedman's Inequality for Matrix Martingales. *Electronic Communications in Probability*, 16, 262-270.

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