

7510678-forecasting-realized-volatility-with-fractional-ornstein-uhlenbeck-processes-a-s

SUMMARY: This paper develops a forecasting framework for realized volatility (RV) using the fractional Ornstein-Uhlenbeck (fOU) process, which is a continuous-time model. The authors estimate the fOU process by approximate Whittle maximum likelihood (AWML) with a modified Paxson spectral kernel, and couple the estimates with a truncated causal best linear predictor (fOU-BLUP) for direct multi-step means. The study compares fOU-BLUP against HAR, free-d ARFIMA, and GARCH models on a panel of liquid high-frequency equity assets. The results show that fOU-BLUP enters the 90% model confidence set against HAR, free-d ARFIMA, and GARCH at the horizons studied, with the clearest MSE gains away from the most heavily mined mega-cap names. The authors also provide an out-of-sample map from residual-scale forecasts to level-scale variance, so that MSE and Patton QLIKE can be read jointly while accounting for loss asymmetry and exponential transformation curvature.

METHOD: The authors use a panel of liquid high-frequency equity assets as the empirical test bed. They estimate the fOU process by approximate Whittle maximum likelihood (AWML) with a modified Paxson spectral kernel. The estimates are coupled with a truncated causal best linear predictor (fOU-BLUP) for direct multi-step means. The study uses a rolling causal horse race design to compare fOU-BLUP against HAR, free-d ARFIMA, and GARCH models.

KEY FINDINGS:

- fOU-BLUP enters the 90% model confidence set against HAR, free-d ARFIMA, and GARCH at the horizons studied.
- fOU-BLUP provides the clearest MSE gains away from the most heavily mined mega-cap names.
- The residual-to-level map links linear L2 optimality on log-scale residuals to Patton QLIKE evaluation on the original realized-volatility level.
- The study provides an out-of-sample map from residual-scale forecasts to level-scale variance.

IMPLICATION FOR TRADING: The findings suggest that encoding spectral roughness in a discrete linear predictor can improve the accuracy of real-time volatility forecasting. This approach could be used by execution desks to schedule child orders and by risk management desks to construct return-density and Value-at-Risk calculations. The study also highlights the importance of using a fully causal out-of-sample design for forecasting, as opposed to regional case studies.

Forecasting Realized Volatility with Fractional Ornstein–Uhlenbeck Processes: A Spectral BLUP Approach

Yifan Ye^a, Zheqi Fan^{b,c,*}

^a*Faculty of Business and Management, Beijing Normal-Hong Kong Baptist University, Zhuhai, China*

^b*Division of Emerging Interdisciplinary Areas, Hong Kong University of Science and Technology, Hong Kong SAR*

^c*Thrust of Financial Technology, Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China*

Abstract

This paper develops a continuous-time forecasting framework for realized volatility using the fractional Ornstein–Uhlenbeck (fOU) process. To overcome the computational burden of exact likelihoods and to bridge continuous dynamics with discrete forecasting, we estimate the fOU by approximate Whittle maximum likelihood (AWML) with a modified Paxson spectral kernel and couple the estimates with a truncated causal best linear predictor (fOU-BLUP) for direct multi-step means. We also spell out the out-of-sample map from residual-scale forecasts to level-scale variance, so that MSE and Patton QLIKE can be read jointly while accounting for loss asymmetry and exponential transformation curvature. On a panel of liquid high-frequency equity assets used as a demanding empirical test bed, rolling causal horse races show that fOU-BLUP enters the 90% model confidence set against HAR, free- d ARFIMA, and GARCH at the horizons we study, with the clearest MSE gains away from the most heavily mined mega-cap names. The results underscore the value of encoding spectral roughness in a discrete linear predictor for real-time volatility forecasting.

Keywords: Fractional Ornstein–Uhlenbeck, Approximate Whittle likelihood, Spectral forecasting, Best linear predictor, Model confidence set, Realized volatility, Trading volume

1 Introduction

Can Chinese equity volume and realized volatility be forecast more accurately once their roughness is built into the predictor? Volume is not a side series: in classical market microstructure, price impact rises with a trader’s participation rate in market volume (Kyle, 1985), and volume itself is a central proxy for liquidity and the intensity of risk

*Corresponding author.

Email address: zheqi.fan@connect.ust.hk (Zheqi Fan)

sharing (Campbell et al., 1993). Recent work elevates that object to a first-class forecast target: Goyenko et al. (2024) show that stock-level volume is highly predictable and that better volume forecasts map into economically large “trading volume alpha” once implementation costs enter the portfolio problem. Execution desks therefore need day-ahead and week-ahead volume paths to schedule child orders. On the risk-management side, realized-volatility forecasts are a standard input to return-density construction, Value-at-Risk, and related capital and margin calculations (Andersen et al., 2003), and comparing volatility measures through an explicit risk-management lens is itself established practice (Brownlees and Gallo, 2010). The same desks that clear overnight and weekly risk also care about expected tape intensity: under-forecasting volume understates market capacity and overstates temporary impact, while over-forecasting volume misaligns hedge schedules and liquidity buffers with the tape that actually arrives. Better forecasts of both objects are therefore operational inputs for A-share execution, hedging, and risk books, and for ETF issuers and authorized participants, before they are an econometric horse race. Workhorse tools such as HAR regressions (Corsi, 2009), ARFIMA, and short-memory ARMA are designed for long memory or persistence, not for the rough, anti-persistent paths that recent continuous-time work associates with equity information flow (Gatheral et al., 2018; Bennedsen et al., 2022; Wang et al., 2023; Shi et al., 2024).

Our research question is whether an AWML-estimated fractional Ornstein–Uhlenbeck best linear predictor (fOU-BLUP) improves day-, week-, and month-ahead forecasts of equity volume and realized volatility relative to HAR, free- d ARFIMA, and short-memory benchmarks under a fully causal out-of-sample design. We answer that question on a demanding high-frequency equity panel used as an empirical stress test rather than as a regional case study.

To the best of our knowledge, this paper makes the following threefold methodological contribution to the volatility forecasting and financial econometrics literature. First, we bridge continuous-time fractional Ornstein–Uhlenbeck dynamics and discrete-time causal forecasting by constructing a truncated best linear predictor that operates on the continuous autocovariance kernel implied by the aliased spectral density, so multi-step means need not rest on an Euler discretization of the SDE. Relative to the day/week/month averages in HAR-RV (Corsi, 2009), the same kernel supplies a smooth memory filter driven by continuous parameters (H, κ) rather than a heuristic piecewise staircase. Second, we deploy the modified Paxson approximate Whittle maximum likelihood (AWML) engine of Shi et al. (2024) so that monthly rolling re-estimation remains computationally feasible for a large out-of-sample horse race, bypassing the $O(N^3)$ bottleneck of exact

time-domain fractional Gaussian likelihoods on dense covariance matrices. Third, we spell out the residual-to-level map that links linear L_2 optimality on log-scale residuals to Patton QLIKE evaluation on the original realized-volatility level, making explicit the Jensen curvature of the exponential transform and why a median-unbiased level forecast need not coincide with residual-scale MSE rankings. Table I summarizes the design contrast.

Table I: Design contrast with the closest related papers. Columns record what is forecast, on which tape, which dynamics are admitted, and how out-of-sample claims are disciplined.

Paper	Tape	Series	Dynamics in play	OOS discipline
Corsi (2009)	U.S.	RV	HAR vs	Multi-horizon MSE
Shi and Yu (2023)	U.S.	RV	long-memory rivals ARFIMA: $d > 0$ vs near-unit anti-persistence	horse race Whittle/finite- sample ID; OOS accuracy
Li et al. (2025)	U.S.	RV + volume	Long memory $\approx$ rough (weak ID)	ID-robust sets; often neither rejected
Wang et al. (2023)	U.S.	RV	fOU	Modeling and forecasts
Shi et al. (2024)	U.S.	Volume	fOU (AWML)	Primarily in-sample H
Goyenko et al. (2024)	U.S.	Volume	Predictability maps	OOS $\rightarrow$ trading-cost alpha
This paper	China A-share	Volume + RV	fOU-BLUP vs HAR and free- d ARFIMA	Causal MA; rolling MCS (MSE; QLIKE for RV)

In sample, Chinese volume and RV are rough. On five liquid index ETFs, AWML on detrended log volume yields $\hat{H} \in [0.10, 0.15]$. For the Huatai-Pinebridge CSI 300 ETF (510300), the two-sided full-sample estimate is $\hat{H}_{\text{AWML}} = 0.1197$ (0.0059). On the CSI 300 cash index, five-minute realized volatility over the same years gives 0.1225 (0.0052). SSE 50, CSI 500, ChiNext, and the Shanghai Composite sit in 0.12–0.14 (Tables VII–VIII and Figure 7). Every interval $\hat{H} + 2$ s.e. lies well below one half. Mean reversion is slow: $\exp(-\hat{\kappa}\Delta)$ is 0.988–0.996 on both objects. Volume and volatility agree.

Out of sample, the rough predictor helps on statistical losses. Under a rolling five-year window, monthly AWML refits, and a *causal* trailing two-year moving average (so that forecasts do not peek at future volume or RV), fOU-BLUP enters the 90% model confidence set of [Hansen et al. \(2011\)](#) for all five ETFs at the one-day horizon and reduces

MSE by roughly one quarter to two fifths relative to MA(22) (Table IX and Figure 8). ARFIMA and a Corsi-style HAR on volume are close competitors; AR(1), ARMA(1,1), and MA(22) are eliminated. On index RV, fOU-BLUP and ARFIMA sit in the MCS with the standard HAR-RV of Corsi (2009) on large-cap names and improve relative MSE on mid- and small-cap indices; GARCH(1,1) mapped from ETF returns does not enter the MCS (Table X).

The economic premise is the mixture-distribution hypothesis of Clark (1973), Tauchen and Pitts (1983), and Andersen (1996): volume and volatility are two readings of the same information flow. If that flow is rough, both Chinese series should produce $H < 1/2$, and predictors that respect roughness should improve density-free point forecasts. Mainland China is a natural laboratory: the market is large enough to matter for global forecasting practice, institutionally different enough to stress-test whether roughness is a U.S. tape artifact, and operationally relevant for users who clear risk on Shanghai and Shenzhen clocks (Section 4.1). Three institutional facts change the implementation. First, the A-share continuous auction runs 09:30–11:30 and 13:00–15:00, so five-minute RV has 48 bins, not 78. Second, a typical year has about 242 sessions, so $\Delta = 1/242$ and the two-year moving average uses $L = 484$. Third, T+1, price limits, suspensions, and ST designations interrupt the tape.

We use daily ETF share volume and five-minute index realized volatility from standard Chinese market data vendors. A simulation study at a known design point recovers H with much smaller RMSE than the change-of-frequency (CoF) start (Wang et al., 2023), which we keep only as a start and benchmark.

We contribute to three strands of literature. Each strand already contains strong forecasting or modeling evidence; our contribution is to couple an explicitly rough continuous-time predictor with a causal out-of-sample discipline that those literatures do not jointly impose, using the empirical panel as a stress test of that synthesis.

The first strand is realized-volatility forecasting. High-frequency realized measures turned latent variance into an observable forecasting object (Andersen et al., 2003), and five-minute RV remains a hard-to-beat sampling default across asset classes (Liu et al., 2015). The workhorse multi-horizon specification is the heterogeneous autoregressive model of Corsi (2009), typically raced against long-memory ARFIMA and short-memory rivals under MSE and related losses, with imperfect-proxy theory guiding which losses are robust when the RV target is noisy (Patton, 2011; Brownlees and Gallo, 2010). A parallel continuous-time literature argues that volatility paths are often rough ($H < 1/2$) rather than classically long-memory fractional (Gatheral et al., 2018; Bennedsen et al.,

2022), and that ARFIMA $d > 0$ can be hard to separate from near-unit anti-persistence in finite samples (Shi and Yu, 2023; Li et al., 2025). Relative to Corsi (2009), we retain HAR as the named RV benchmark but read its day/week/month averages as a heuristic piecewise filter against the smooth memory kernel implied by (H, κ) , and test whether the rough continuous-time predictor co-survives in a causal horse race. Relative to Shi and Yu (2023) and Li et al. (2025), we keep free- d ARFIMA as a rival and ask whether the rough predictor survives out of sample with HAR. Relative to Wang et al. (2023), we add matched volume, enforce causal detrending, and race the BLUP against HAR and free- d ARFIMA. The evaluation language stays statistical (MSE, and QLIKE for RV) and MCS-based (Hansen et al., 2011), which matches this journal’s forecast-comparison tradition.

The second strand is trading-volume predictability. Classical microstructure treats volume as the scale of risk sharing and the denominator of a trader’s participation rate in price impact (Kyle, 1985; Campbell et al., 1993), while mixture-distribution theory links volume and volatility as two readings of the same information flow (Clark, 1973; Tauchen and Pitts, 1983; Andersen, 1996). Empirically, volume is highly predictable at the stock level, and better volume forecasts map into economically large after-cost “trading volume alpha” once implementation costs enter the portfolio problem (Goyenko et al., 2024). Continuous-time rough dynamics have already been taken to U.S. equity volume with AWML estimates of H well below one half (Shi et al., 2024), but that evidence is primarily in-sample on a U.S. tape. Relative to Goyenko et al. (2024), we evaluate Chinese ETF volume on MSE forecast losses rather than mapping predictability into after-cost portfolio alpha. Relative to Shi et al. (2024), we apply AWML-fOU to both Chinese ETF volume and index RV and discipline claims with a causal rolling MCS. The practical stake is execution-relevant: day-ahead and week-ahead volume paths are inputs to child-order schedules, not only descriptive microstructure.

The third strand is Chinese equity-market dynamics under rough continuous-time models. Mainland A-shares differ from the U.S. tape on clocks and frictions that change every forecasting implementation: a lunch break that cuts five-minute RV to 48 bins, about 242 sessions per year so that $\Delta = 1/242$, and T+1, price limits, and retail intensity that can amplify high-frequency reversals. Existing China-focused volume–RV forecasting work documents predictability and cross-series links on the domestic tape (Liu et al., 2023), but does not embed an explicitly rough fOU predictor or the AWML engine. Rough-volatility and rough-volume evidence is still concentrated on developed Western markets (Gatheral et al., 2018; Shi et al., 2024; Wang et al., 2023). Relative to that body

of work, we ask whether roughness and forecast gains survive Shanghai and Shenzhen clocks rather than only U.S. equity series, and whether matched Chinese volume and RV agree as two readings of one information flow. Table I summarizes the design contrast with the closest papers; the Chinese application is the geography of the claim, not a change of econometric object.

The empirical panel is deliberately chosen as a rigorous testing ground rather than a convenience sample: high retail participation, distinct trading frictions, and strong time-varying volatility generate roughness and clustering that stress HAR-style predictors, while avoiding the over-fitting risks associated with well-mined U.S. benchmark indices.

Section 2 records the fOU, AWML estimation, and a simulation check. Section 3 develops the causal BLUP predictor and the residual-to-level loss map. Section 4 describes the asset universe and volatility measurement. Section 5 reports in-sample estimates. Section 6 evaluates out-of-sample MSE, QLIKE, and MCS. Section 7 concludes.

2 The fractional Ornstein–Uhlenbeck framework

This section records the continuous-time object, its aliased spectral density, and the AWML estimator that the forecasting engine uses. A small simulation study checks the discrete implementation in the rough region before any market data are used.

2.1 The fractional Ornstein–Uhlenbeck process

We use one continuous-time object for both forecast targets. Under the mixture-distribution hypothesis, volume and volatility are two readings of the same information flow (Clark, 1973; Andersen, 1996); under rough volatility, that flow is anti-persistent with $H < 1/2$ (Gatheral et al., 2018; Bennedsen et al., 2022), a property that also improves VIX term-structure fit relative to jump-augmented Heston benchmarks (Ye et al., 2026). The stationary fractional Ornstein–Uhlenbeck (fOU) nests Brownian mean reversion ($H = 1/2$), fractional Brownian motion ($\kappa = 0$), and the rough, mean-reverting case that recent Whittle work estimates on U.S. equity volume (Shi et al., 2024) and on RV (Wang et al., 2023). The empirical application changes only the sampling clock Δ and the construction of the observed path x ; the spectral density below is the same object that later feeds the best linear predictor in Section 3.

Let B^H be a two-sided fractional Brownian motion with Hurst index $H \in (0, 1)$ (Mandelbrot and Van Ness, 1968): a centered Gaussian process with

$$\mathbb{E}[B_t^H B_s^H] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t - s|^{2H}). \quad (2.1)$$

The increment process is fractional Gaussian noise. Covariances of that noise are not summable when $H > 1/2$ (long memory) and sum to zero when $H < 1/2$ (anti-persistence). Paths are locally Hölder of order $H - \varepsilon$. The case $H = 1/2$ is ordinary Brownian motion.

The stationary fOU (Cheridito et al., 2003; Hu and Nualart, 2010) is the solution of

$$dY_t = \kappa(\mu - Y_t) dt + \sigma dB_t^H, \quad t \in \mathbb{R}, \quad (2.2)$$

with $\kappa > 0$ and with Y_0 drawn from the invariant law $\mathcal{N}(\mu, \sigma^2 \kappa^{-2H} H \Gamma(2H))$. The pathwise solution is

$$Y_t = e^{-\kappa t} Y_0 + \mu(1 - e^{-\kappa t}) + \sigma \int_0^t e^{-\kappa(t-s)} dB_s^H. \quad (2.3)$$

When $H = 1/2$ the process is the classical Ornstein–Uhlenbeck diffusion. When $\kappa = 0$ it collapses to fractional Brownian motion. We keep $\kappa > 0$ throughout, as in Shi et al. (2024). The parameter space is the compact set in Assumption 2.1.

Assumption 2.1. $(\kappa, H, \sigma) \in \Theta$, a compact subset of $\mathbb{R}_+ \times (0, 1) \times \mathbb{R}_+$.

The continuous-time spectral density is (Garnier and Sølna, 2018)

$$f(\omega; \kappa, H, \sigma^2) = \frac{\sigma^2}{2\pi} c_H \frac{|\omega|^{1-2H}}{\kappa^2 + \omega^2}, \quad c_H = \Gamma(2H + 1) \sin(\pi H), \quad \omega \in \mathbb{R}. \quad (2.4)$$

Lemma 2.1 of Shi et al. (2024) states that this density identifies (κ, H, σ) on Θ . Empirical series are observed at a fixed interval $\Delta > 0$. Write $X_j = Y_{j\Delta}$ and n for the sample length, so that the span is $n\Delta$. Aliasing produces

$$f_\Delta(\lambda; \kappa, H, \sigma^2) = \frac{\sigma^2}{2\pi} c_H \Delta^{2H} \sum_{j=-\infty}^{\infty} \frac{|\lambda + 2\pi j|^{1-2H}}{(\kappa\Delta)^2 + (\lambda + 2\pi j)^2}, \quad \lambda \in (0, 2\pi). \quad (2.5)$$

The infinite sum is the computational problem. A plain truncation at $\pm K$ has uniform error of order $O(K^{-2H})$, which is usable when $H > 1/2$ but not in the rough region that equity volume and rough-volatility estimates occupy: as $H \rightarrow 0$ the truncation error stays $O(1)$ no matter how large K is (Shi et al., 2024). Rough paths are exactly the region that matters (Gatheral et al., 2018; Bennedsen et al., 2022; Fukasawa et al., 2022; Bolko et al., 2023; Wang et al., 2023). For the equity calendar used below we set $\Delta = 1/242$ (about 242 sessions per year); the same aliased density is evaluated at that Δ , so the estimator and the later BLUP inherit that clock rather than a U.S. 1/250 convention.

2.2 Approximate Whittle estimation

With the continuous-time object in hand, we record the AWML objects from [Shi et al. \(2024, Sections 3–4\)](#) that the Chinese application calls. We do not reprove the theorems. Two implementation choices matter for forecasting: the modified Paxson kernel at $K = 200$, which remains accurate when H is near zero, and the use of CoF only as a start; inference and the OOS predictor use AWML.

2.2.1 Modified Paxson spectral density

Split the aliased sum at a cutoff K and write

$$S(\lambda) = \sum_{j=1}^{\infty} a_{1,j}(\lambda) + a_0(\lambda) + \sum_{j=1}^{\infty} a_{2,j}(\lambda), \quad (2.6)$$

with

$$a_0(\lambda) = \frac{|\lambda|^{1-2H}}{\theta^2 + \lambda^2}, \quad a_{1,j}(\lambda) = \frac{(2\pi j - \lambda)^{1-2H}}{\theta^2 + (2\pi j - \lambda)^2}, \quad a_{2,j}(\lambda) = \frac{(2\pi j + \lambda)^{1-2H}}{\theta^2 + (2\pi j + \lambda)^2}, \quad (2.7)$$

and $\theta = \kappa\Delta$. The summands $a_{1,j}$ and $a_{2,j}$ are eventually monotone in j once $K > \theta/(2\pi)$ ([Shi et al., 2024, Lemma 3.1](#)). For daily data with $\Delta \approx 1/242$ and κ of the size estimated on U.S. volume, that threshold is just above one. The tails then admit closed-form upper and lower bounds. Averaging those bounds yields the modified Paxson density

$$\begin{aligned} \tilde{f}_{\Delta}(\lambda) = \frac{\sigma^2}{2\pi} c_H \Delta^{2H} & \left[\sum_{j=1}^K a_{1,j}(\lambda) + \sum_{j=1}^K a_{2,j}(\lambda) + a_0(\lambda) \right. \\ & \left. + \frac{1}{2}Q_{11} + \frac{1}{2}Q_{12} + \frac{1}{2}Q_{21} + \frac{1}{2}Q_{22} \right], \end{aligned} \quad (2.8)$$

where the four Q terms are the explicit tail integrals

$$\begin{aligned} Q_{11} &= \frac{1}{4\pi H} (2\pi K + \lambda)^{-2H}, \\ Q_{12} &= \frac{1}{4\pi H} (2\pi(K+1) + \lambda)^{-2H} - \frac{\theta^2}{2\pi(2H+2)} (2\pi(K+1) + \lambda)^{-2H-2}. \end{aligned} \quad (2.9)$$

and Q_{21} , Q_{22} are the same expressions with λ replaced by $-\lambda$ in the obvious way. Theorem 3.1 of [Shi et al. \(2024\)](#) gives a uniform error $O(K^{-2H-1})$. At H near zero that is $O(K^{-1})$, which still vanishes. The truncation method does not. We follow [Shi et al. \(2024\)](#) and set $K = 200$.

2.2.2 The AWML criterion

Let $I_n(\lambda)$ be the periodogram of the demeaned sample $\{x_j\}_{j=1}^n$,

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{j=1}^n x_j e^{-ij\lambda} \right|^2. \quad (2.10)$$

When μ is unknown we replace x_j by $x_j - \bar{x}$. The sample mean satisfies the rate required by [Dahlhaus \(1989\)](#) and [Lieberman et al. \(2012\)](#) for both the long-memory and the anti-persistent cases ([Shi et al., 2024](#), Section 4.1.2). The Whittle contrast ([Whittle, 1953](#); [Fox and Taqqu, 1986](#)) with f_Δ replaced by $\tilde{f}_\Delta$ is

$$\hat{\vartheta} := (\hat{H}, \hat{\kappa}) = \arg \min_{H, \kappa} \left\{ \log \left(\frac{1}{\pi} \int_0^\pi \frac{I_n(\lambda)}{\tilde{s}(\lambda; H, \kappa)} d\lambda \right) + \frac{1}{\pi} \int_0^\pi \log \tilde{s}(\lambda; H, \kappa) d\lambda \right\}, \quad (2.11)$$

where $\tilde{s}$ is the scale-free Paxson kernel (the term in brackets in (2.8)). The scale is recovered in closed form,

$$\hat{\sigma}^2 = \frac{1}{\pi} \int_0^\pi \frac{I_n(\lambda)}{\tilde{s}(\lambda; \hat{H}, \hat{\kappa})} d\lambda. \quad (2.12)$$

Equations (2.11)–(2.12) are the concentrated AWML criterion. Numerical minimization is Nelder–Mead, started at the CoF estimates of Section 2.2.4.

When $H > 1/2$ the density blows up at the origin. The criterion then integrates only on $[2\pi/n, \pi]$ and adds the two correction terms of [Shi et al. \(2024, Lemma 4.1\)](#). U.S. volume estimates all lie below $1/2$, so the correction is a safeguard rather than the leading case. We keep it, because a Chinese name can in principle come out persistent.

2.2.3 Asymptotics and standard errors

Theorem 2.2 ([Shi et al. 2024](#), Theorems 4.1–4.2). *Let the true value $\vartheta_0 = (\kappa_0, H_0, \sigma_0)$ lie in the interior of Θ . As $n \rightarrow \infty$ with Δ fixed, the Whittle estimator based on f_Δ is consistent and*

$$\sqrt{n}(\hat{\vartheta}_W - \vartheta_0) \rightarrow \mathcal{N}(0, \Gamma_0^{-1}), \quad \Gamma_0 = \frac{1}{4\pi} \int_0^{2\pi} (\partial \log f_\Delta(\lambda)) (\partial \log f_\Delta(\lambda))' d\lambda. \quad (2.13)$$

If in addition $\sqrt{n} K^{-2H_0-1} \rightarrow 0$, the AWML estimator is asymptotically equivalent to the Whittle estimator.

The information matrix again contains infinite sums. Lemma 4.2 of the same paper approximates the derivatives $\partial_H S$ and $\partial_\kappa S$ by Paxson tails. We report those analytic

standard errors for AWML. For CoF we use the long-span / shrinking- Δ formulae of Wang et al. (2023, Theorem 4.1). Two features of Theorem 2.2 matter for the empirical panel. The limiting variance is continuous in $H \in (0, 1)$. The CoF and composite-likelihood limits are not (Wang et al., 2023; Bennedsen et al., 2026). And the condition on K is trivial at $K = 200$ for any H that the U.S. estimates have produced.

2.2.4 Change-of-frequency start

Let $\Delta_1 x_j = x_{j+1} - x_j$ and $\Delta_2 x_j = x_{j+2} - x_j$, and write $\Delta_{11}x$ and $\Delta_{22}x$ for the corresponding second differences. The CoF estimator of Wang et al. (2023) is

$$\hat{H}_{\text{CoF}} = \frac{1}{2} \log_2 \left(\frac{\sum_j (\Delta_{22} x_j)^2}{\sum_j (\Delta_{11} x_j)^2} \right), \quad (2.14)$$

with $\hat{\sigma}_{\text{CoF}}^2$ recovered from the second-difference quadratic variation and $\hat{\kappa}_{\text{CoF}}$ from the sample variance of x after $\hat{H}$ and $\hat{\sigma}^2$ are plugged in. CoF is $\sqrt{n}$ -consistent for H under an in-fill plus long-span scheme, but its variance deteriorates as $H \rightarrow 0$ and the theory for κ depends on the true H . We use it only as a start and as a published benchmark, not as the preferred estimator.

2.3 Simulation study

Before taking AWML to equity volume and RV, we check that the discrete implementation recovers a known design point in the rough region that matters for those series. The exercise is a simulation study of estimator recovery, not a forecast horse race: the data-generating process is an Euler–Maruyama fOU driven by Davies–Harte fractional Brownian motion, with $(\kappa, \mu, \sigma, H) = (10, 0, 1, 0.2)$, $\Delta = 1/250$, eight infill steps, and a five-year span, the same neighbourhood used to validate AWML against CoF in the continuous-time literature (Shi et al., 2024; Wang et al., 2023). The concentrated Whittle contrast (2.11) is evaluated on the Fourier frequencies $\lambda_j = 2\pi j/n$, $j = 1, \dots, \lfloor (n-1)/2 \rfloor$, with the Paxson kernel at $K = 200$. Minimization is Nelder–Mead, started at CoF. This is the same discrete criterion that produces Tables VII–VIII. The continuous-frequency contrast in Shi et al. (2024) is asymptotically equivalent; the discrete Fourier form is what we take to the empirical panel.

On a single validation path of length $n = 501$, CoF returns $(\hat{H}, \hat{\kappa}, \hat{\sigma}) = (0.1402, 2.26, 0.71)$. Discrete AWML returns $(0.1995, 8.21, 0.99)$, against the truth $(0.2, 10, 1)$. Table II reports fifty independent replications at $n = 1250$: columns are CoF and AWML for (H, κ, σ) ; rows are the single-path check, Monte Carlo means, standard deviations, and RMSE

against truth. AWML bias for H is -0.004 and RMSE is 0.017 , against CoF RMSE 0.040 . The gain is larger for κ : AWML RMSE 3.01 against CoF 7.20 . Both estimators understate σ by about 0.11 . The design therefore prefers AWML for roughness and mean reversion on paths with H near 0.2 ; CoF remains useful as a start value. We do not interpret Chinese $\hat{\sigma}$ as a structural volatility of information flow (scale is absorbed by the log transform and demeaning in Section 4), so the simulation claim that matters for forecasting is the recovery of H and κ .

Table II: Simulation study at $(H, \kappa, \sigma) = (0.2, 10, 1)$, $\Delta = 1/250$, $n = 1250$, 50 replications. RMSE is against the true value. Single-path row: $n = 501$ validation draw.

	CoF			AWML		
	H	κ	σ	H	κ	σ
Single path ($n = 501$)	0.1402	2.26	0.71	0.1995	8.21	0.99
MC mean	0.1928	10.56	0.89	0.1961	10.39	0.89
MC std. dev.	0.0397	7.26	0.20	0.0164	3.01	0.07
MC RMSE	0.0399	7.20	0.22	0.0167	3.01	0.13

3 Causal out-of-sample forecasting via BLUP

In-sample roughness is necessary but not sufficient for forecast usefulness. This section turns the AWML engine into a real-time predictor: a causal trailing detrend, a truncated spectral BLUP for direct multi-step means, named rivals, and the residual-to-level map used for QLIKE.

The forecasting exercise keeps the AWML engine and changes only what a real-time user could know. In-sample roughness tables may use a two-sided moving average; the horse race uses a causal trailing average of the same length.

Windows and targets. Estimation uses a rolling window of 1210 trading days (≈ 5 years). Parameters are re-estimated once per calendar month; within the month the frozen parameters produce daily forecasts. Monthly refits balance the cost of re-optimizing the Whittle contrast against parameter drift on a rough path. The out-of-sample calendar starts at 2017-01-01 when the rolling window is available, and otherwise at the first origin with a full window after listing and causal detrend (so 510300 begins in mid-2019). Table III lists each series' OOS start, end, and number of day-ahead origins ($n_{\text{oos}} \approx 1,150\text{--}1,920$). Horizons are $h \in \{1, 5, 22\}$. For $h = 1$ the target is the next day's residual; for $h \in \{5, 22\}$ it is the average of the next h residuals (direct multi-step, as in

the realized-volatility forecasting tradition of [Andersen et al. \(2003\)](#) and [Corsi \(2009\)](#)). Direct multi-step avoids iterating a one-step map under a misspecified short-memory dynamics.

Table III: Out-of-sample evaluation window. Rolling estimation uses 1210 trading days with monthly parameter refits; n_{oos} is the number of forecast origins at $h = 1$.

Object	Ticker	OOS start	OOS end	n_{oos}
Volume	510300.SH	2019-05-13	2024-11-29	1350
Volume	510050.SH	2017-01-03	2024-11-29	1921
Volume	510500.SH	2020-03-03	2024-11-29	1153
Volume	159915.SZ	2018-11-26	2024-11-29	1458
Volume	159919.SZ	2019-05-15	2024-11-29	1348
RV	000300	2017-01-03	2024-11-29	1921
RV	000016	2017-01-03	2024-11-29	1921
RV	000905	2017-01-03	2024-11-29	1921
RV	399006	2017-05-24	2024-11-29	1828
RV	000001	2017-01-03	2024-11-29	1921

From AWML to the BLUP. Monthly AWML delivers $\hat{\theta} = (\hat{H}, \hat{\kappa}, \hat{\sigma}^2)$ on the causal residual. The modified Paxson density $\tilde{f}_{\Delta}(\lambda; \hat{\theta})$ in (2.8) is the spectral density of that residual on the daily grid. The corresponding autocovariance on lags $k = 0, 1, \dots$ is the Fourier dual

$$\gamma(k; \hat{\theta}) = \int_0^{2\pi} e^{ik\lambda} \tilde{f}_{\Delta}(\lambda; \hat{\theta}) d\lambda, \quad (3.1)$$

evaluated numerically on the Fourier grid used for Whittle. The truncated best linear predictor of length $L = 60$ targets the same direct multi-step mean as HAR and AR: $\bar{x}_{t,h} = \frac{1}{h} \sum_{j=1}^h x_{t+j}$ (so $h = 1$ is the usual one-step case). With history $x_t, \dots, x_{t-L+1}$,

$$\hat{\bar{x}}_{t,h|t} = \sum_{\ell=1}^L w_{\ell,h} x_{t+1-\ell}, \quad \Gamma_L(\hat{\theta}) w_h = \bar{\gamma}_h(\hat{\theta}), \quad (3.2)$$

where Γ_L has entries $\gamma(|i - j|; \hat{\theta})$ and

$$\bar{\gamma}_h(\hat{\theta}) = \frac{1}{h} \sum_{j=1}^h (\gamma(j; \hat{\theta}), \dots, \gamma(j + L - 1; \hat{\theta}))^{\top}. \quad (3.3)$$

When $h = 1$, $\bar{\gamma}_1$ reduces to $(\gamma(1), \dots, \gamma(L))^{\top}$. Truncation at two months keeps the predictor operational while capturing the slow mean reversion implied by small $\hat{\kappa}\Delta$. Relative

MSE for representative series is essentially unchanged for $L \in \{20, 40, 60, 80\}$ (Online Appendix B), so we keep $L = 60$ as the main-text default. The same map runs for volume and for $\log \sqrt{\text{RV}}$ residuals. Relative to free- d ARFIMA, which iterates a discrete path and then averages, the BLUP encodes the continuous-time rough spectral shape through $(\hat{H}, \hat{\kappa})$ in one linear projection on the mean target.

HAR features. For both volume and RV residuals, the HAR competitor uses day/week/month averages in the spirit of Corsi (2009):

$$x_t^{(d)} = x_t, \quad x_t^{(w)} = \frac{1}{5} \sum_{j=0}^4 x_{t-j}, \quad x_t^{(m)} = \frac{1}{22} \sum_{j=0}^{21} x_{t-j}, \quad (3.4)$$

and a direct multi-step OLS fit of $\text{mean}(x_{t+1}, \dots, x_{t+h})$ on $(1, x_t^{(d)}, x_t^{(w)}, x_t^{(m)})$ over the rolling window.

Volume models. (i) **fOU-BLUP** as in (3.2). (ii) **ARFIMA**(1, d , 0) with d from a Geweke–Porter–Hudak regression, free of the fOU restriction $d = \hat{H} - 1/2$. We allow $d \in (-0.5, 0.5)$, so the semi-parametric estimator can land in the anti-persistent region as well as classical long memory (Hurvich and Ray, 1995; Shi and Yu, 2023). Empirically, rolling GPH estimates on causally detrended volume concentrate near the upper stationary boundary (asset-level means about 0.48–0.49), so the horse race pits an explicitly rough continuous-time predictor against a freely estimated long-memory discrete competitor rather than against a forced $d = \hat{H} - 1/2$ plug-in—a contrast that is hard to settle by in-sample fit alone (Li et al., 2025). (iii) **HAR-Volume** with features (3.4). (iv)–(vi) Direct **AR**(1), conditional least-squares **ARMA**(1,1), and **MA**(22) as short-memory and pure smoother lower bounds. If roughness or long memory matters, these three should lose on MSE and exit the MCS.

RV models. The same fOU-BLUP (3.2) and ARFIMA run on causally detrended $\log \sqrt{\text{RV}}$. Rolling GPH d on the RV residual is likewise positive on average (pool mean about 0.47), again selecting classical long memory rather than the rough-link value $d = \hat{H} - 1/2 < 0$. Competitors are the standard **HAR-RV** of Corsi (2009) with features (3.4), the named benchmark for RV forecasting since Andersen et al. (2003), and **GARCH**(1,1) on matched ETF daily returns, variance-targeted and scaled to the RV level on the estimation window. Related work shows that option-implied jumps and rough-Heston spot volatility can improve HAR-style RV forecasts (Yu et al., 2025; Fan

et al., 2025); our horse race stays on the cash-index RV path and does not use an options panel. GARCH is a deliberately harsh benchmark: it uses a different information set (daily closes) and must be projected onto index RV, so failure is expected if high-frequency roughness is the relevant object. We do not report ad hoc fractional-HAR variants in the main text; free d in ARFIMA already covers the “fractional differencing then HAR” idea without a mechanical plug-in of $\hat{H} - 1/2$.

Losses and inference. Volume and RV headline losses are MSE (and MAE) on the residual scale. Equal predictive ability is summarized by the 90% model confidence set of Hansen et al. (2011) with a stationary block bootstrap. For multi-step horizons ($h = 5, 22$), the block length is $l = \max(5, h)$ so that overlapping direct targets, which induce an $\text{MA}(h - 1)$ structure in the forecast errors, do not understate the sampling variability of loss differentials. Stars in the tables mark membership in the 90% MCS ($\alpha = 0.10$). Because five-minute RV is a noisy proxy for latent integrated variance, MSE alone is not robust to proxy noise (Patton, 2011); as a robustness check we also evaluate Patton’s QLIKE loss

$$\text{QLIKE}_{t+h} = \frac{\text{RV}_{t+h}}{\widehat{\text{RV}}_{t+h}} - \log\left(\frac{\text{RV}_{t+h}}{\widehat{\text{RV}}_{t+h}}\right) - 1. \quad (3.5)$$

MSE and MAE are computed directly on the causally detrended residual x_t^{rv} . QLIKE is scale-dependent and requires strictly positive variance levels, so residual forecasts are mapped back to the RV scale before evaluating (3.5):

$$\widehat{\text{RV}}_{t+h} = \exp\left(2\left[\widehat{x}_{t,h|t}^{\text{rv}} + \hat{\mu} + \text{MA}_{484,t}(\log \text{RVol})\right]\right), \quad (3.6)$$

where $\text{MA}_{484,t}$ is the causal trailing two-year average of $\log \text{RVol}$ dated at the forecast origin (no look-ahead), and $\hat{\mu}$ is the sample mean of the causally detrended residual path used to center that residual before forecasting. Because the trailing MA already removes the slow level of activity, $\hat{\mu}$ is a near-zero location constant rather than a look-ahead trend filter; residual-scale MSE and level-scale QLIKE share the same residual definition, and we do not re-estimate a separate demean inside the rolling horse race. Under a Gaussian forecast error on the log scale, (3.6) delivers a median-unbiased variance forecast: by Jensen’s inequality, $\mathbb{E}[e^Z] \geq e^{\mathbb{E}[Z]}$, so the plain exponential is not the conditional mean. We deliberately refrain from an ad hoc half-variance convexity correction $\frac{1}{2}\hat{\sigma}_e^2$, which would inject auxiliary estimation noise into every origin. The resulting mild

convexity underprediction, together with QLIKE’s asymmetric penalty on understated spikes, further helps explain why level-scale GARCH can look relatively stronger under QLIKE at $h = 22$ (Online Appendix A). GARCH forecasts are already produced as scaled variance levels and enter (3.5) directly. The associated 90% MCS uses the same block length. Relative QLIKE versus HAR-RV and QLIKE-MCS inclusions are reported in Online Appendix A (Table XI). Pairwise Diebold–Mariano statistics versus fOU-BLUP are reported with Harvey–Leybourne–Newbold small-sample adjustment in Online Appendix B (Diebold and Mariano, 1995; Harvey et al., 1997); the main text leads with MCS because the design is a multi-model horse race, not a single pairwise contest.

4 Data and volatility measurement

Section 2 specifies the fOU and shows that discrete AWML recovers H and κ in a rough design; Section 3 records the causal predictor. We now construct the paths that enter estimation and the horse race.

4.1 Asset universe and market context

The empirical test bed is a panel of liquid Chinese equity-index realized-volatility series and matched index ETF volume paths. Mainland continuous auctions run 09:30–11:30 and 13:00–15:00, so five-minute RV has 48 bins rather than a U.S.-style continuous session; a typical year has about 242 sessions, which fixes $\Delta = 1/242$ and a two-year moving average of length $L = 484$. T+1 settlement, daily price limits, and heavy retail participation generate abrupt high-frequency reversals that stress standard short-memory and HAR-style predictors.¹

The Chinese equity market provides an exceptionally rigorous testing ground for continuous-time rough and long-memory models. Characterized by high retail participation, distinct trading frictions, and substantial time-varying volatility dynamics, it displays pronounced roughness and clustering that severely challenge standard discrete-time linear specifications such as HAR-RV. Demonstrating out-of-sample predictability in such a vibrant and volatile environment provides strong empirical validation for the fOU-BLUP framework, while avoiding the over-fitting risks associated with well-mined U.S. benchmark indices. Daily ETF share volume and five-minute cash-index realized volatility are drawn from commercial Chinese market databases. The universe covers CSI 300, SSE 50, CSI 500, ChiNext, and the Shanghai Composite, with liquid ETFs paired to the cash indices

¹We study mainland exchanges only; Hong Kong is outside the sample.

(Table IV). The sample runs from 2010 through 2024 where listings and complete five-minute coverage permit (we retain a day only if at least 40 of the 48 continuous-auction bins are available).

Table IV: ETF volume series and matched cash-index RV series. Volume is ETF share volume; RV is five-minute index realized variance.

ETF (volume)	Index (RV)	Basket
510300.SH	000300.SH	CSI 300
510050.SH	000016.SH	SSE 50
510500.SH	000905.SH	CSI 500
159915.SZ	399006.SZ	ChiNext
159919.SZ	000300.SH	CSI 300 (Harvest ETF)

Table V: Level-scale summary statistics, 2010-01-04–2024-12-31 (where listed). Volume mean/median are share volume in millions. RVol is five-minute realized volatility in percent-return units. Ann. vol is $\text{RVol} \times \sqrt{242}$ in percent for display only; estimation and forecasting use log RVol after MA detrend (Table VI).

Series	Ticker	n	Mean	Median	Ann. vol (%)
<i>Panel A. ETF share volume (millions)</i>					
CSI 300 ETF	510300.SH	3064	5.1	3.6	–
SSE 50 ETF	510050.SH	3643	6.0	4.8	–
CSI 500 ETF	510500.SH	2867	1.5	1.1	–
ChiNext ETF	159915.SZ	3172	4.4	3.0	–
CSI 300 ETF (Harvest)	159919.SZ	3062	1.3	0.9	–
<i>Panel B. Index RVol ($\sqrt{\text{RV}}$)</i>					
CSI 300	000300.SH	3641	1.002	0.873	15.6
SSE 50	000016.SH	3641	1.014	0.877	15.8
CSI 500	000905.SH	3641	1.083	0.935	16.8
ChiNext	399006.SZ	3542	1.315	1.155	20.5
Shanghai Composite	000001.SH	3641	0.902	0.763	14.0

4.2 Summary statistics

Table V reports level-scale summaries. ETF share volume averages a few million shares per day. Index realized variance is the sum of squared five-minute *percent* returns (units $\%^2$), so typical $\text{RVol}_t = \sqrt{\text{RV}_t}$ values are of order one; the table’s ann. vol column, $\text{RVol}_t \times \sqrt{242}$ in percent, puts the same series on a familiar scale (about 14–21% on average) for display only. Estimation and forecasting never annualize: they use $\log \text{RVol}_t$

after the MA transform below, so a constant change of units cancels. We follow the industry and academic default of five-minute RV (Liu et al., 2015) and use vendor five-minute quadratic variation rather than rebuilding bins from higher-frequency stamps.

4.3 Volume measurement and filtering

The volume transform is the additive-in-logs detrend that isolates residual dynamics for Whittle. Let V_t be share volume. Form

$$y_t^{\text{vol}} = \log V_t - \text{MA}_L(\log V_t), \quad L = 484, \quad x_t^{\text{vol}} = y_t^{\text{vol}} - \overline{y^{\text{vol}}}. \quad (4.1)$$

The moving average is two-sided and centered for the in-sample roughness tables. Drop the first 241 and last 241 observations so that every remaining y_t uses a full window. Then demean. That is the additive-in-logs detrend of Andersen (1996) and of Shi et al. (2024, Section 5): it removes a slowly moving level of activity so that the Whittle contrast targets the residual dynamics that determine H and κ . For out-of-sample forecasts we replace the centered MA with a causal trailing MA of the same length, so the residual at origin t uses only dates $\leq t$. A constant unit change—lots versus shares—cancels under the log. We do not winsorize. The 2015 crash stays in the path, because a forecast user on the A-share tape cannot pretend those days away, and because roughness claims that vanish after crisis trimming would not be useful for desks.

Figure 1 plots $\log V_t$ for 510300 together with the two-sided MA that defines the in-sample residual; the residual path itself appears later as Figure 4.

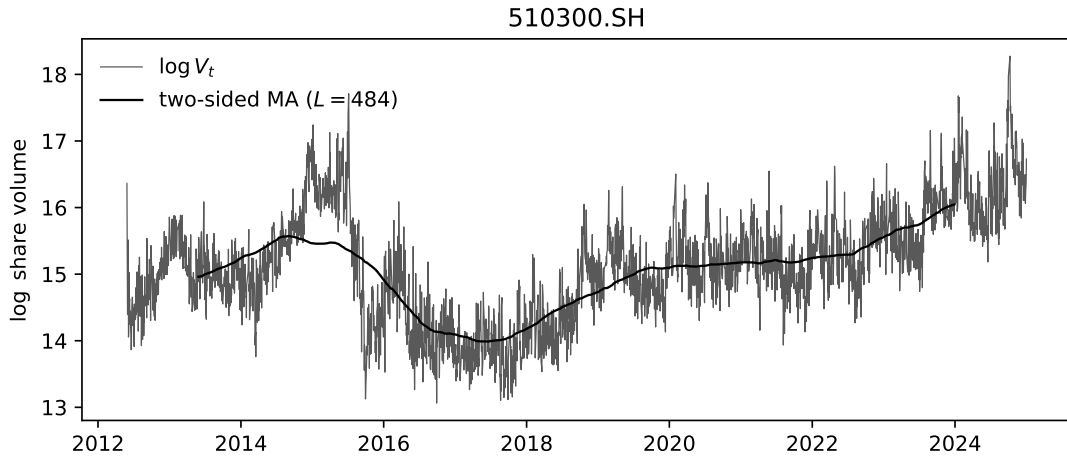


Figure 1: 510300.SH: log share volume and the two-sided two-year MA used in the in-sample detrend.

4.4 Realized volatility construction and filtering

The RV path is built to match the volume residual construction under the same A-share calendar. Let RV_t denote five-minute realized variance in percent-return units and set $RVol_t = \sqrt{RV_t}$. Apply the same transform for in-sample tables (two-sided MA) and the causal trailing MA for forecasting,

$$y_t^{rv} = \log RVol_t - \text{MA}_{484}(\log RVol_t), \quad x_t^{rv} = y_t^{rv} - \overline{y^{rv}}. \quad (4.2)$$

Overnight and lunch-break returns are excluded from RV_t . The log transform is invariant to a constant scale of RV_t , so H and κ do not depend on whether variance is stored in percent-squared units or a proportional transform.

Figure 2 shows CSI 300 on an annualized-percent display scale for readability; the model still uses $\log RVol$ after (4.2).

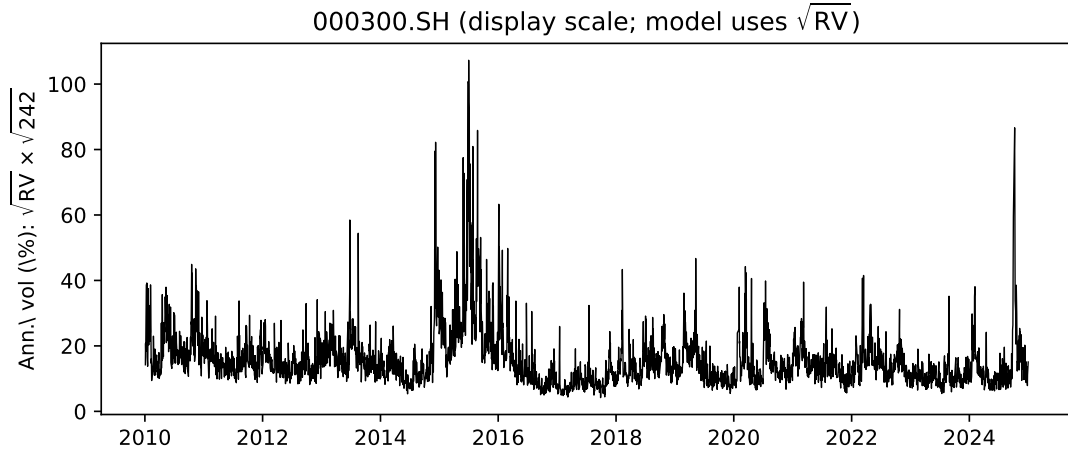


Figure 2: 000300.SH: annualized display $RVol_t \times \sqrt{242}$ in percent. Estimation uses $\log RVol_t$ after MA detrend, not this annualized series.

$$RV_t = \sum_{i=1}^{48} r_{t,i}^2, \quad RVol_t = \sqrt{RV_t}. \quad (4.3)$$

Equation (4.3) is the daily realized-variance object used throughout. Figure 3 records the A-share clock that delivers 48 five-minute bins: twenty-four morning and twenty-four afternoon, with the lunch break excluded. Relative to a U.S. five-minute RV with 78 bins over a continuous session, the A-share design has fewer intraday observations per day and a hard midday gap; both features are absorbed into the daily $RVol$ series before the

fOU is estimated, so the continuous-time $\Delta = 1/242$ remains the correct daily sampling interval for Whittle.

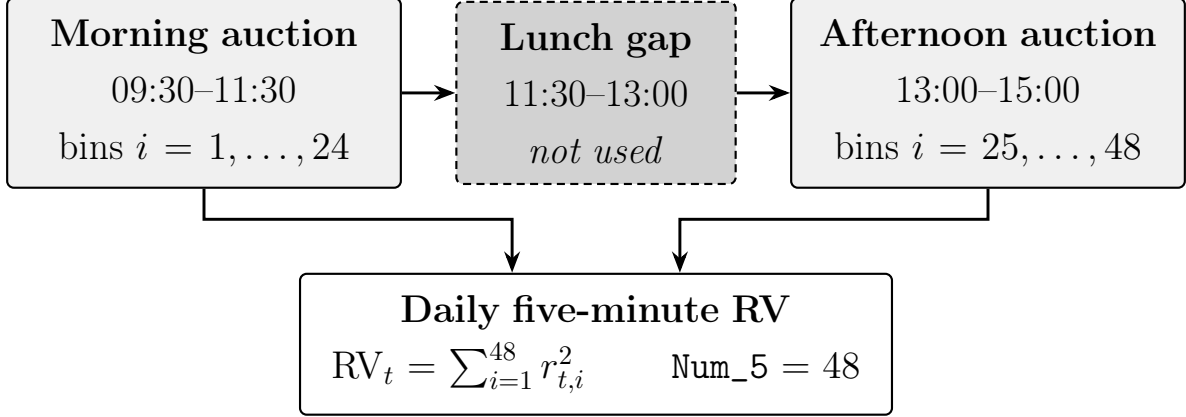


Figure 3: A-share session clock for five-minute RV. Morning and afternoon continuous auctions each contribute 24 bins; the 11:30–13:00 lunch break is excluded, so a full day has 48 bins (versus 78 on a continuous U.S. session).

5 In-sample estimation results

With the residual paths of Section 4 in hand, we ask whether Chinese ETF volume and matched index RV are rough under AWML, and whether the two readings of information flow agree. Out-of-sample predictors are deferred to Sections 3–6; this section is only the in-sample roughness claim.

Estimation follows the AWML engine of Section 2.2 on the Chinese calendar: $\Delta = 1/242$, Paxson cutoff $K = 200$, and the two-sided MA detrend of Section 4. Nelder–Mead minimization of the concentrated discrete Whittle contrast starts at the CoF estimates of Section 2.2.4; when CoF is inadmissible ($H \notin (0, 1)$ or non-positive κ or σ^2), the start is $(H, \kappa) = (0.2, 5)$. Standard errors for (H, κ) come from the numerical Hessian. CoF columns in the tables are reported only as a published start and benchmark, never as the preferred estimate: the simulation study in Section 2.3 already prefers AWML for roughness and mean reversion near $H = 0.2$. For each name the scientific claim is whether $\hat{H}_{\text{AWML}} + 2 \text{ s.e.}$ lies below one half. We do not treat CoF as a test of roughness, and we do not estimate a joint volume–volatility system. The implied daily AR coefficient $\widehat{\text{AR}} = \exp(-\hat{\kappa}_{\text{AWML}}/242)$ is reported only to translate $\hat{\kappa}$ into a familiar one-day persistence scale.

Tables VI–VIII use the ETF volume and index RVol panels constructed in Section 4. Dates are the first and last observations of x after the 241-day trim at each end, so the

estimation window ends in December 2023 even though raw series run through 2024-12-31. The claim we write is whether Chinese $\hat{H}_{\text{AWML}}$ sits well below one half, in the rough band documented for U.S. equity volume by [Shi et al. \(2024\)](#) and for rough volatility by [Gatheral et al. \(2018\)](#), and whether matched volume and RV agree. We do not run a formal China–U.S. equality test; the comparison is qualitative and external-validity minded.

5.1 Sample description

Before reporting $\hat{H}$, Table VI lists each series, ticker, the start and end of the demeaned residual x_t after the two-year MA trim, the effective length n , and the sample standard deviation of x_t . Level-scale means are in Table V. Volume samples begin after listing plus trim (510300 from 2013-05-28, $n = 2580$; 510050 is the longest volume panel at $n = 3159$). Index RVol panels are aligned on 2011–2023 with $n \approx 3157$ except ChiNext ($n = 3058$). Residual sd is about 0.49–0.63 for volume and 0.36–0.39 for RVol: volume residuals are noisier in levels, but both objects are long enough for Whittle estimation at daily Δ . The table is the sampling frame for every subsequent $\hat{H}$ and for the OOS windows in Section 3.

Table VI: Sample description of the detrended series x_t . Start and End are after dropping 241 days at each end of the two-year moving average.

Series	Ticker	Start	End	n	sd of x_t
Volume, CSI 300 ETF	510300.SH	2013-05-28	2023-12-29	2580	0.5006
Volume, SSE 50 ETF	510050.SH	2011-01-04	2023-12-29	3159	0.4926
Volume, CSI 500 ETF	510500.SH	2014-03-19	2023-12-29	2383	0.5334
Volume, ChiNext ETF	159915.SZ	2012-12-07	2023-12-28	2684	0.5387
Volume, CSI 300 ETF (Harvest)	159919.SZ	2013-05-29	2023-12-28	2574	0.6262
RVol, CSI 300 index	000300.SH	2011-01-04	2023-12-29	3157	0.3655
RVol, SSE 50 index	000016.SH	2011-01-04	2023-12-29	3157	0.3766
RVol, CSI 500 index	000905.SH	2011-01-04	2023-12-29	3157	0.3886
RVol, ChiNext index	399006.SZ	2011-06-02	2023-12-29	3058	0.3569
RVol, Shanghai Composite	000001.SH	2011-01-04	2023-12-29	3157	0.3724

5.2 ETF volume estimates

The first roughness claim is on ETF share volume, the information-flow proxy that execution desks observe directly. Table VII reports CoF and AWML estimates of (H, κ, σ) on detrended log ETF volume. AWML standard errors (parentheses) come from the nu-

merical Hessian of the concentrated discrete Whittle contrast. All five AWML $\hat{H}$ lie between 0.099 and 0.154. Each two-standard-error interval lies below 0.13 except ChiNext (0.1540, still below 0.17). The implied daily AR coefficients $\exp(-\hat{\kappa}/242)$ are 0.9926, 0.9922, 0.9884, 0.9915, and 0.9962: mean reversion is slow, as on U.S. volume in Shi et al. (2024). CoF is reported as a start and published benchmark only; where CoF and AWML disagree (e.g. 510500, 159915), we follow the simulation preference for AWML. The scientific reading is that A-share ETF volume is rough, not Brownian: $H = 1/2$ is rejected at conventional two-standard-error bands for every name.

Figure 4 plots the demeaned residual x_t for 510300 after (4.1)—the series that enters the periodogram. The path wanders with persistent low-frequency swings and sharp bursts around 2015 and 2020; nothing in the figure suggests a short-memory AR(1) world. Figure 7(a) puts the five AWML $\hat{H}$ with two-standard-error bars next to the Brownian line $H = 1/2$ and the SPY point 0.1631 from Shi et al. (2024): every Chinese interval sits well left of one half, and most sit near or below the SPY benchmark.

Table VII: fOU estimates on Chinese ETF log volume. AWML standard errors in parentheses are from the numerical Hessian of the concentrated discrete Whittle contrast.

Ticker	CoF			AWML		
	H	κ	σ	H	κ	σ
510300.SH	0.1693	5.5687	1.0021	0.1197 (0.0059)	1.7982 (0.4930)	0.7648
510050.SH	0.1239	1.7797	0.7856	0.1183 (0.0052)	1.8960 (0.4461)	0.7634
510500.SH	0.1803	12.7186	1.2644	0.1165 (0.0065)	2.8172 (0.6487)	0.8961
159915.SZ	0.2217	7.0708	1.2487	0.1540 (0.0065)	2.0559 (0.5098)	0.8600
159919.SZ	0.1564	4.7513	1.1938	0.0987 (0.0053)	0.9134 (0.3972)	0.8739

5.3 Realized-volatility estimates

The second claim is on cash-index RV under the same AWML protocol, so that volume and volatility can be compared as two readings of one flow. Table VIII repeats the CoF/AWML layout for five-minute index RVol after (4.2). All five AWML $\hat{H}$ lie between 0.120 and 0.141, again far below one half. Matched pairs agree: CSI 300 volume 0.1197 versus CSI 300 RV 0.1225; SSE 50 volume 0.1183 versus SSE 50 RV 0.1196. Under the

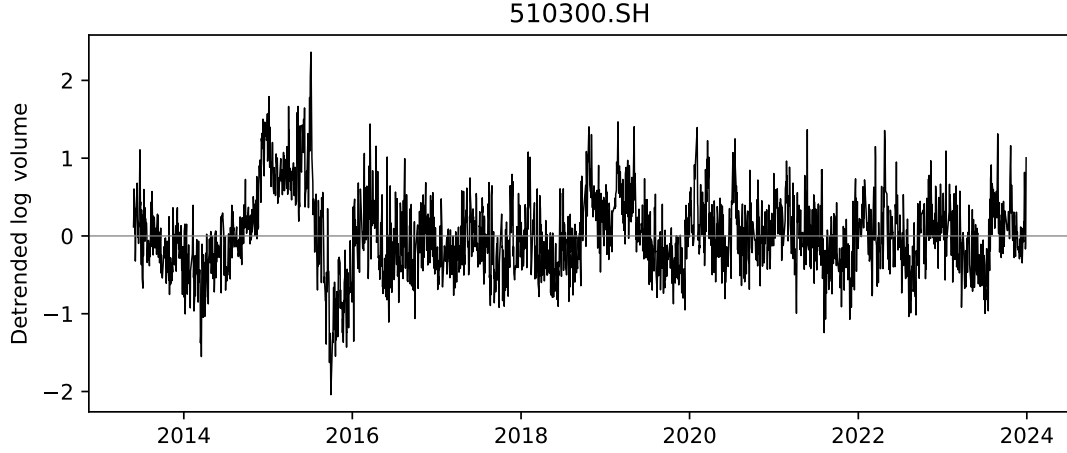


Figure 4: Detrended log share volume of 510300.SH after (4.1). This is the series that enters CoF and AWML.

mixture-distribution hypothesis (Clark, 1973; Andersen, 1996), volume and volatility are two readings of the same rough information flow; the Chinese estimates are consistent with that premise and with the rough-volatility evidence of Gatheral et al. (2018) and Bennedsen et al. (2022) and the fOU–RV modelling of Wang et al. (2023), now on A-share clocks.

Figure 6 shows the CSI 300 demeaned log RVol path: the same crash and COVID spikes visible in volume, with a quieter mid-sample. Figure 5 aligns the CSI 300 volume and RVol residuals on a common calendar: the two paths share crisis bursts without being identical, which is the visual counterpart of the mixture premise before the matched $\hat{H}$ comparison in Figure 7(b). Figure 7(b) stacks matched volume and RV estimates for CSI 300, SSE 50, and ChiNext; pairs sit on top of each other within sampling error, which is the in-sample half of the paper’s contribution in Table I—roughness is not an artifact of studying volume alone.

The estimated Hurst parameters for Chinese equity series ($\hat{H} \approx 0.10\text{--}0.15$) are consistently lower than those reported for U.S. markets ($\hat{H} \approx 0.16$ for SPY volume in Shi et al., 2024). This heightened degree of roughness—stronger anti-persistence at high frequencies—can be rationalized by institutional frictions on the A-share tape. Under the T+1 trading rule, domestic investors facing adverse intraday shocks cannot liquidate the same day, so hedging and rebalancing demand accumulates and tends to erupt at subsequent openings. Coupled with heavy retail participation and short-horizon overreaction, local fluctuations experience abrupt reversals, which is consistent with rougher

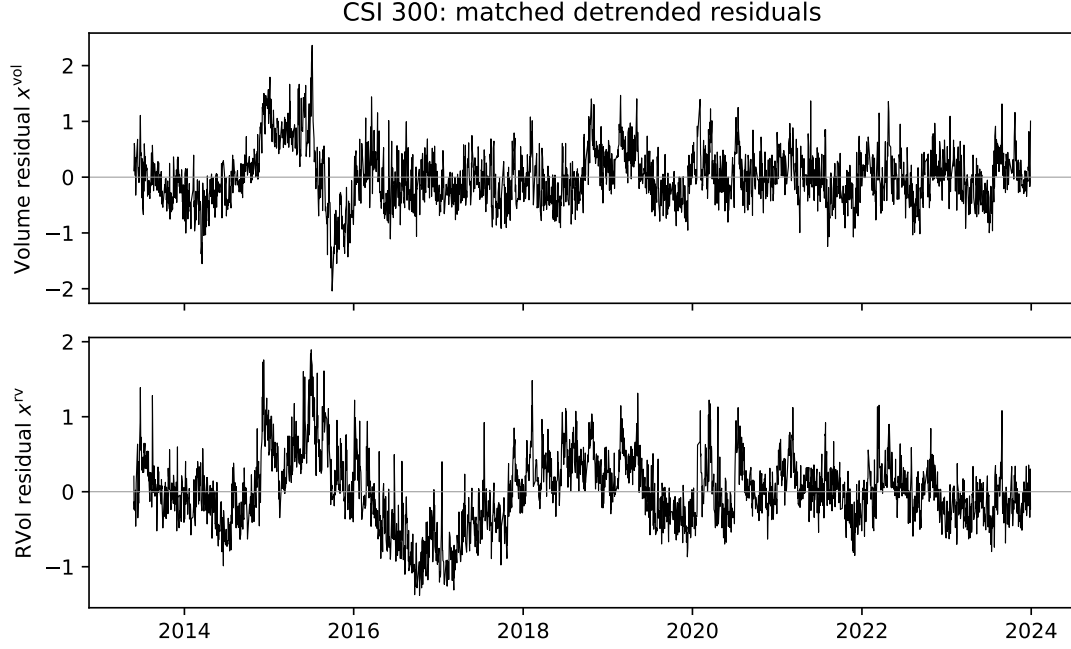


Figure 5: CSI 300 matched residuals after the two-year MA detrend: ETF volume (510300, top) and cash-index RVol (000300, bottom) on common dates.

Table VIII: fOU estimates on five-minute realized volatility after (4.2). AWML standard errors are from the numerical Hessian.

Ticker	CoF			AWML		
	H	κ	σ	H	κ	σ
000300.SH	0.0872	0.2944	0.4829	0.1225 (0.0052)	2.3683 (0.4944)	0.5855
000016.SH	0.0694	0.0667	0.4559	0.1196 (0.0051)	2.3864 (0.4948)	0.5985
000905.SH	0.1965	10.2136	0.9208	0.1413 (0.0062)	2.9741 (0.5666)	0.6823
399006.SZ	0.1644	5.5930	0.7085	0.1248 (0.0056)	2.1783 (0.4869)	0.5719
000001.SH	0.1023	0.6032	0.5221	0.1332 (0.0055)	2.6778 (0.5274)	0.6196

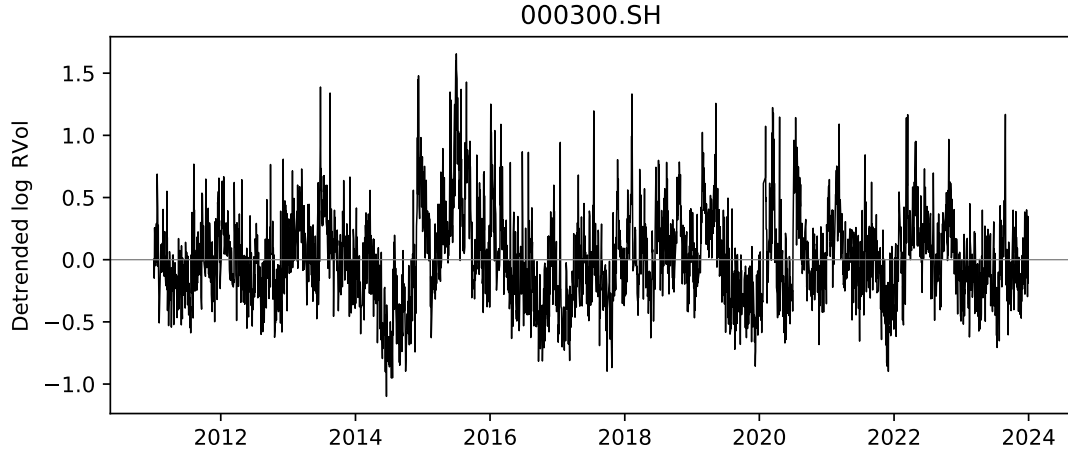


Figure 6: Detrended log realized volatility of 000300.SH after (4.2).

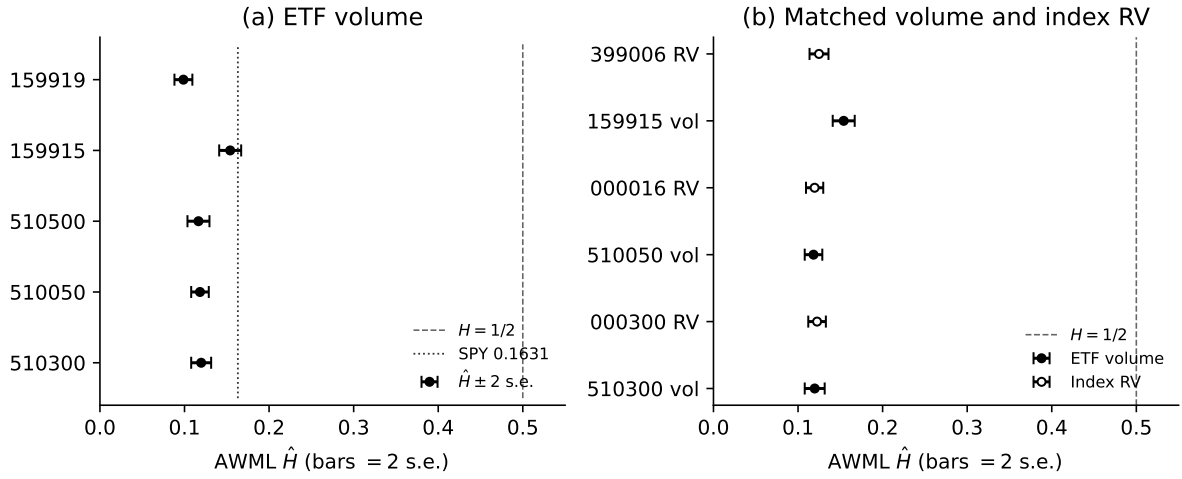


Figure 7: AWML $\hat{H}$ with two-standard-error bars. (a) Five Chinese index ETF volumes; dashed line $H = 1/2$, dotted line SPY estimate 0.1631 of Shi et al. (2024). (b) Matched ETF volume and cash-index RV for CSI 300, SSE 50, and ChiNext; dashed line $H = 1/2$.

continuous-time paths. We treat this as an economic reading of the estimates, not as a formal identification of $T+1$ or retail share as causes of H .

6 Out-of-sample forecast evaluation

With the causal design of Section 3 fixed, we report MSE and MCS membership for volume and then for RV. Table IX reports relative MSE versus MA(22) for ETF volume across $h = 1, 5, 22$. Columns are the six predictors; a star marks membership in the 90% MCS. At $h = 1$, fOU-BLUP is in the MCS for all five ETFs and cuts MSE by 21–39% relative to MA(22) (relative MSE 0.61–0.79). Relative to the industrial multi-horizon benchmark HAR-Volume, the gains are smaller—on the order of one percentage point of relative MSE (e.g., 0.7379 versus 0.7462 on 510300 at $h = 1$)—but they are enough for fOU-BLUP to survive the MCS whenever HAR does, and often when HAR does not. ARFIMA is almost as strong (MCS inclusion rate 1.00 across names at $h = 1$). HAR-Volume enters often but less systematically (inclusion rate 0.60 at $h = 1$). AR(1), ARMA(1,1), and MA(22) are never in the MCS: short-memory and pure smoothers are eliminated once roughness or flexible long memory is admitted. At longer horizons the ranking is stable for fOU-BLUP and ARFIMA on most names; on 510500 at $h = 5, 22$ ARFIMA edges fOU-BLUP into a shared or ARFIMA-led MCS. That pattern is consistent with free GPH d selecting classical long memory: multi-step averaging washes out high-frequency roughness and leaves a smoother persistence component that a positive- d ARFIMA can exploit. The practical reading is that a rough (or freely fractional) predictor belongs in the day-ahead volume toolkit alongside HAR, that MA(22)—still common as a desk smoother—is not competitive under MSE, and that the headline win versus MA(22) should not be confused with the more modest but MCS-relevant edge versus HAR-Volume.

Figure 8 illustrates the 510300 panel. Bars are relative MSE versus MA(22) by horizon; the dashed line is the MA(22) benchmark at one. fOU, ARFIMA, and HAR sit well below one at every h , while AR(1) and ARMA rise toward or above one at $h = 22$. The figure is the visual counterpart of the first block of Table IX: on the CSI 300 ETF that plays the SPY role for A-shares, encoding roughness (or free d) pays in point-forecast accuracy for execution-style horizons.

Table X and Figure 9 turn to the RV horse race, with relative MSE versus HAR-RV. fOU-BLUP is in the 90% MCS for every index and every horizon (inclusion rate 1.00). ARFIMA co-survives at most horizons. On CSI 300 and SSE 50 the relative MSE versus

Table IX: Out-of-sample volume forecast losses (MSE) and MCS p -values. Relative MSE versus MA(22). HAR-Volume is the industrial multi-horizon benchmark; compare columns fOU-BLUP and HAR-Vol for the high-spec contrast. * denotes inclusion in the 90% model confidence set ($\alpha = 0.10$) using the stationary block bootstrap.

Asset	h	fOU-BLUP	ARFIMA	HAR-Vol	AR(1)	ARMA(1,1)	MA(22)
510300.SH	1	0.7379*	0.7378*	0.7462*	0.8697	0.7943	1.0000
	5	0.7149*	0.7170*	0.7346	0.9954	0.9477	1.0000
	22	0.7003*	0.7009*	0.7364*	1.0271	1.0818	1.0000
510050.SH	1	0.7636*	0.7625*	0.7735	0.9289	0.8226	1.0000
	5	0.7354*	0.7314*	0.7515	1.0856	0.9619	1.0000
	22	0.7650*	0.7588*	0.7884*	1.2211	1.4618	1.0000
510500.SH	1	0.7366*	0.7322*	0.7467	0.8576	0.7830	1.0000
	5	0.7533	0.7410*	0.7776	1.0595	0.9758	1.0000
	22	0.7521	0.7230*	0.8063	1.2705	1.5725	1.0000
159915.SZ	1	0.6127*	0.6202*	0.6149*	0.7172	0.6577	1.0000
	5	0.6258*	0.6618	0.6259*	0.8589	0.8491	1.0000
	22	0.7028*	0.7663	0.6880*	0.8606	1.3903	1.0000
159919.SZ	1	0.7861*	0.7830*	0.7909*	0.9877	0.8726	1.0000
	5	0.8037*	0.7951*	0.7973*	1.3004	1.2831	1.0000
	22	0.8829*	0.8374*	0.8321*	1.4440	2.9801	1.0000
MCS90 rate	1	1.00	1.00	0.60	0.00	0.00	0.00
MCS90 rate	5	0.80	0.80	0.40	0.00	0.00	0.00
MCS90 rate	22	0.80	0.80	0.80	0.00	0.00	0.00

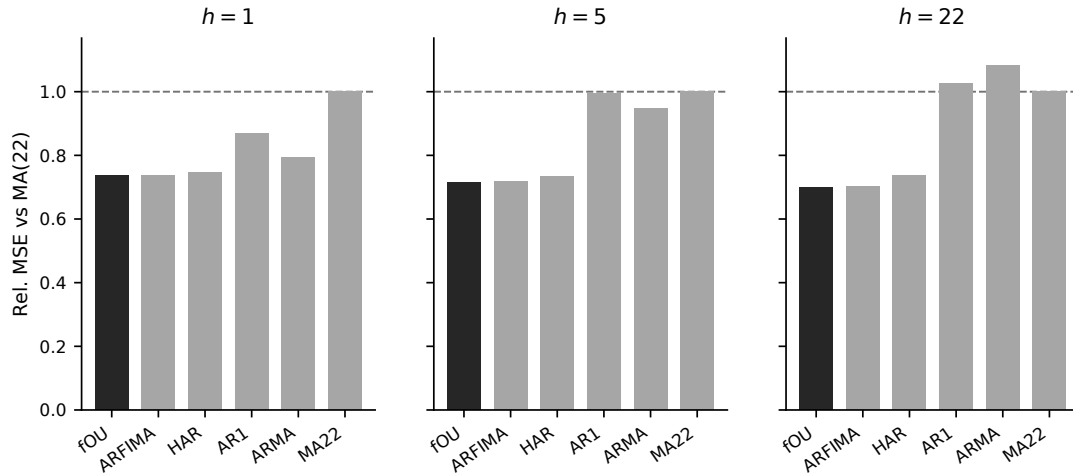


Figure 8: 510300.SH volume: relative OOS MSE versus MA(22) by horizon. Dark bar: fOU-BLUP. Dashed line: MA(22) benchmark.

Table X: Out-of-sample RV forecast losses. Relative MSE versus HAR-RV. * denotes inclusion in the 90% model confidence set ($\alpha = 0.10$) using the stationary block bootstrap.

Asset	h	fOU-BLUP	HAR-RV	ARFIMA	GARCH
000300	1	1.0015*	1.0000*	0.9962*	1.5750
	5	0.9773*	1.0000*	0.9900*	1.7347
	22	0.9622*	1.0000*	0.9573*	1.6370
000016	1	1.0073*	1.0000*	0.9977*	1.6802
	5	0.9867*	1.0000*	0.9856*	2.0026
	22	0.9888*	1.0000*	0.9531*	1.9620
000905	1	0.9845*	1.0000	0.9884*	2.5642
	5	0.9732*	1.0000	0.9765*	3.9784
	22	0.9388*	1.0000	0.9326*	3.9180
399006	1	0.9792*	1.0000	0.9826*	1.4256
	5	0.9371*	1.0000	0.9492	1.4966
	22	0.8889*	1.0000	0.9079*	1.3061
000001	1	0.9973*	1.0000*	0.9974*	2.0765
	5	0.9680*	1.0000	1.0092	2.3326
	22	0.9447*	1.0000	0.9627*	2.1229
MCS90 rate	1	1.00	0.60	1.00	0.00
MCS90 rate	5	1.00	0.40	0.60	0.00
MCS90 rate	22	1.00	0.40	1.00	0.00

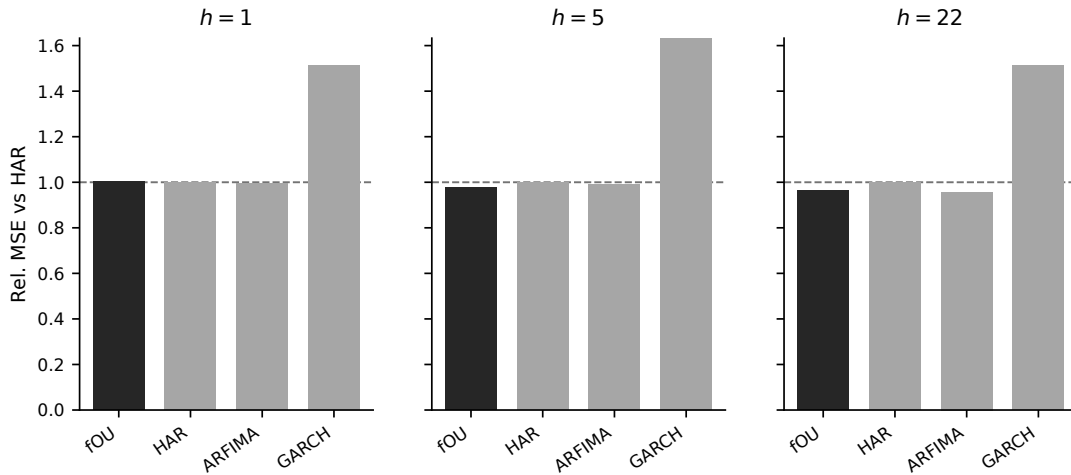


Figure 9: CSI 300 index RV: relative OOS MSE versus HAR-RV by horizon. Dark bar: fOU-BLUP. Dashed line: HAR-RV benchmark.

HAR-RV is close to one (about 0.96–1.01), so the practical message is MCS *co-membership* rather than a large point-loss gap against the Corsi benchmark on large-cap names. On CSI 500 and ChiNext the relative MSE falls more clearly below one (down to about 0.89 at $h = 22$ on ChiNext), so the rough predictor’s edge is sharper where the tape is less blue-chip. GARCH(1,1) never enters the MSE-based MCS: mapping ETF conditional variance into index RV does not pay under this loss. Figure 9 makes the CSI 300 pattern visible—fOU and ARFIMA hug the HAR line—while the table shows where the gain concentrates off the mega-cap slice. Under Patton QLIKE after the level map (3.6) (Online Appendix A), fOU-BLUP remains in the QLIKE MCS at every horizon on four of five indices and at $h = 5, 22$ on all five; relative QLIKE versus HAR-RV stays near one on large-caps and below one on CSI 500 and ChiNext, paralleling the MSE geography. At the monthly horizon the QLIKE ranking for GARCH can look stronger than under MSE—a loss-asymmetry effect documented in Appendix A—so we treat the MSE horse race as the main-text headline and QLIKE as the proxy-noise robustness check required by Patton (2011).

Taken together, the OOS evidence answers the journal question in Table I: once $H < 1/2$ is estimated on Chinese volume and RV, the implied fOU predictor earns a place in the MCS against HAR and ARFIMA, and it dominates short-memory alternatives under MSE. That is a forecast-accuracy claim, not a claim that roughness alone replaces HAR on every large-cap RV path. Cumulative squared forecast-error differentials versus HAR (Online Appendix B) rise steadily on 510300 volume and at $h = 5$ on CSI 300 RV, rather than in a single crisis jump, so those MSE gains are not outlier-driven; at $h = 1$ on CSI 300 RV the cumulative differential stays near zero, matching the near-parity relative MSE against HAR-RV. An equal-weight fOU–HAR combination and Mincer–Zarnowitz encompassing regressions (Mincer and Zarnowitz, 1969; Bates and Granger, 1969) show that the two predictors are complementary on CSI 300 RV and that fOU largely encompasses HAR on 510300 volume (Appendix B). Pairwise Diebold–Mariano statistics with the Harvey–Leybourne–Newbold adjustment (Harvey et al., 1997) are reported in the same appendix; the main text continues to lead with MCS because the design is a multi-model horse race (Diebold and Mariano, 1995).

7 Conclusion

The paper asked whether building roughness into the predictor improves forecasts of Chinese ETF volume and index RV. The in-sample and out-of-sample evidence say yes

on MCS grounds, with the usual caveats on large-cap RV versus HAR. Chinese ETF share volume and Chinese index realized volatility are rough in sample, and that roughness is useful for point forecasts out of sample. AWML places H well below one half on both objects, with the CSI 300 volume/RV pair at 0.120 versus 0.122, in line with the rough band in [Gatheral et al. \(2018\)](#); [Bennedsen et al. \(2022\)](#) and the U.S. volume evidence in [Shi et al. \(2024\)](#). In rolling horse races with causal detrending, the fOU best linear predictor enters the 90% MCS for volume on every ETF at $h = 1$ —cutting MSE by about one fifth to two fifths versus MA(22)—and for RV on every index at every horizon studied, alongside ARFIMA; HAR remains a close competitor on large-cap RV, while short-memory and GARCH benchmarks do not.

Relative to the positioning in Table I, the contribution is a joint China volume–RV application of the AWML–fOU engine under A-share clocks, evaluated as an OOS multi-benchmark MCS horse race rather than as in-sample H alone, and rather than as after-cost portfolio alpha in the sense of [Goyenko et al. \(2024\)](#). For practice, the result tells execution and risk users that an explicitly rough predictor—the AWML-spectrum BLUP in (3.2)—belongs in the day-ahead toolkit next to HAR and ARFIMA; it does not say that HAR is obsolete on every mega-cap RV path.

Boundaries remain. The sample is mainland index ETFs and cash indices under $\Delta = 1/242$ and a causal two-year MA; overnight and lunch returns are outside five-minute RV; we do not estimate a joint volume–volatility system. Extending the horse race to single-name stocks, Stock Connect flows, or economically weighted volume losses is left for further work.

Acknowledgements

Yifan Ye was supported by the Guangdong and Hong Kong Universities “1+1+1” Joint Research Collaboration Scheme (Number R0800021-25) and the Beijing Normal-Hong Kong Baptist University Start Up Research Fund (Number UICR0700136-26).

CRedit authorship contribution statement

Yifan Ye: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. Zheqi Fan: Conceptualization, Writing – review & editing.

References

- Andersen, T.G., 1996. Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance* 51, 169–204.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. *Operational Research Quarterly* 20, 451–468.
- Bennedsen, M., Christensen, K., Christensen, P.K., 2026. To be or not to be: Roughness or long memory in volatility? *arXiv preprint ArXiv:2403.12653*.
- Bennedsen, M., Lunde, A., Pakkanen, M.S., 2022. Decoupling the short- and long-term behavior of stochastic volatility. *Journal of Financial Econometrics* 20, 961–1006. doi:[10.1093/jjfinec/nbaa049](https://doi.org/10.1093/jjfinec/nbaa049).
- Bolko, A.E., Christensen, K., Pakkanen, M.S., Veliyev, B., 2023. A GMM approach to estimate the roughness of stochastic volatility. *Journal of Econometrics* 235, 745–778.
- Brownlees, C.T., Gallo, G.M., 2010. Comparison of volatility measures: A risk management perspective. *Journal of Financial Econometrics* 8, 29–56. doi:[10.1093/jjfinec/nbp009](https://doi.org/10.1093/jjfinec/nbp009).
- Campbell, J.Y., Grossman, S.J., Wang, J., 1993. Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics* 108, 905–939.
- Cheridito, P., Kawaguchi, H., Maejima, M., 2003. Fractional Ornstein–Uhlenbeck processes. *Electronic Journal of Probability* 8, 1–14.
- Clark, P.K., 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41, 135–155.
- Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174–196.
- Dahlhaus, R., 1989. Efficient parameter estimation for self-similar processes. *Annals of Statistics* 17, 1749–1766.
- Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.
- Fan, Z., Wang, M.M., Ye, Y., 2025. On Options-Driven Realized Volatility Forecasting: Information Gains via Rough Volatility Model. Available at SSRN: <https://ssrn.com/abstract=5974814> or <http://dx.doi.org/10.2139/ssrn.5974814>. December 21, 2025.
- Fox, R., Taqqu, M.S., 1986. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* 14, 517–532.
- Fukasawa, M., Takabatake, T., Westphal, R., 2022. Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. *Mathematical Finance* 32, 1086–1132.

- doi:[10.1111/mafi.12354](https://doi.org/10.1111/mafi.12354).
- Garnier, J., Sølna, K., 2018. Option pricing under fast-varying and rough stochastic volatility. *Annals of Finance* 14, 489–516. doi:[10.1007/s10436-018-0325-4](https://doi.org/10.1007/s10436-018-0325-4).
- Gatheral, J., Jaisson, T., Rosenbaum, M., 2018. Volatility is rough. *Quantitative Finance* 18, 933–949.
- Goyenko, R., Kelly, B.T., Moskowitz, T.J., Su, Y., Zhang, C., 2024. Trading volume alpha. Working Paper 33037. National Bureau of Economic Research. doi:[10.3386/w33037](https://doi.org/10.3386/w33037). also SSRN 4802345.
- Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. *Econometrica* 79, 453–497.
- Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13, 281–291.
- Hu, Y., Nualart, D., 2010. Parameter estimation for fractional Ornstein–Uhlenbeck processes. *Statistics and Probability Letters* 80, 1030–1038.
- Hurvich, C.M., Ray, B.K., 1995. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of Time Series Analysis* 16, 17–41.
- Kyle, A.S., 1985. Continuous auctions and insider trading. *Econometrica* 53, 1315–1335.
- Li, J., Phillips, P.C.B., Shi, S., Yu, J., 2025. Weak identification of long memory with implications for volatility modeling. *Review of Financial Studies* 38, 3117–3148. doi:[10.1093/rfs/hhaf022](https://doi.org/10.1093/rfs/hhaf022).
- Lieberman, O., Rosemarin, R., Rousseau, J., 2012. Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes. *Econometric Theory* 28, 457–470. doi:[10.1017/S0266466611000399](https://doi.org/10.1017/S0266466611000399).
- Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics* 187, 293–311.
- Liu, M., Choo, W.C., Lee, C.C., Lee, C.C., 2023. Trading volume and realized volatility forecasting: Evidence from the China stock market. *Journal of Forecasting* 42, 76–100. doi:[10.1002/for.2897](https://doi.org/10.1002/for.2897).
- Mandelbrot, B.B., Van Ness, J.W., 1968. Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10, 422–437.
- Mincer, J., Zarnowitz, V., 1969. The evaluation of economic forecasts, in: Mincer, J. (Ed.), *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. National Bureau of Economic Research, pp. 3–46.
- Patton, A.J., 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246–256.

- Shi, S., Yu, J., 2023. Volatility puzzle: Long memory or antipersistence. *Management Science* 69, 3861–3883. doi:[10.1287/mnsc.2022.4552](https://doi.org/10.1287/mnsc.2022.4552).
- Shi, S., Yu, J., Zhang, C., 2024. On the spectral density of fractional Ornstein–Uhlenbeck processes. *Journal of Econometrics* 245, 105872.
- Tauchen, G.E., Pitts, M., 1983. The price variability-volume relationship on speculative markets. *Econometrica* 51, 485–505.
- Wang, X., Xiao, W., Yu, J., 2023. Modeling and forecasting realized volatility with the fractional Ornstein–Uhlenbeck process. *Journal of Econometrics* 232, 389–415.
- Whittle, P., 1953. Estimation and information in stationary time series. *Arkiv för Matematik* 2, 423–434. doi:[10.1007/BF02590998](https://doi.org/10.1007/BF02590998).
- Ye, Y., Fan, Z., Kwok, Y.K., 2026. VIX term structure in the rough Heston model via Markovian approximation. *Journal of Futures Markets* 46, 799–823. doi:[10.1002/fut.70082](https://doi.org/10.1002/fut.70082).
- Yu, J., Ruan, X., Fan, Z., 2025. Merton (1976) implied jump. *Journal of Economic Dynamics and Control* 180, 105199. doi:[10.1016/j.jedc.2025.105199](https://doi.org/10.1016/j.jedc.2025.105199).

Online Appendix A. QLIKE robustness for RV forecasts

Five-minute realized variance is a noisy proxy for latent integrated variance. Following [Patton \(2011\)](#), we evaluate the QLIKE loss (3.5) on the reconstructed variance level (3.6): residual-scale predictors (fOU-BLUP, HAR-RV, ARFIMA) are mapped by adding back the residual location $\hat{\mu}$ and the causal trailing MA before exponentiating; GARCH forecasts already live on the variance scale. As noted in the main text, the plain exponential is median-unbiased on the log scale and we do not apply a half-variance convexity correction. Table XI reports relative QLIKE versus HAR-RV and marks 90% MCS membership under QLIKE, using the same rolling design and block length $l = \max(5, h)$ as in the main text.

At short horizons ($h = 1, 5$), GARCH(1,1) remains a weak QLIKE competitor and is typically outside the MCS. At $h = 22$, however, its relative QLIKE can fall below one on several indices and MCS inclusion rises. That pattern is expected under asymmetric QLIKE, which heavily penalizes underpredictions of volatility spikes: variance-targeted GARCH reverts smoothly toward the long-run unconditional variance and supplies a conservative monthly baseline, whereas high-frequency path-tracking models can incur large asymmetric penalties when local spikes are understated on a month-ahead average target. The main-text claim therefore continues to rest on the MSE horse race, where GARCH never enters the MCS; the QLIKE appendix is a proxy-noise robustness check, not a second headline ranking.

Across the panel, fOU-BLUP remains in the QLIKE MCS for every index at $h = 5$ and $h = 22$ (inclusion rate 1.00) and for four of five indices at $h = 1$ (inclusion 0.80); on CSI 300 at $h = 1$ the relative QLIKE is slightly above one and HAR/ARFIMA occupy the set. Relative QLIKE versus HAR-RV stays near one on large-caps and below one on CSI 500 and ChiNext, so the geography still parallels Table X.

Table XI: Online Appendix A. Out-of-sample RV forecast losses under Patton (2011) QLIKE. Relative QLIKE versus HAR-RV. * denotes inclusion in the 90% model confidence set under QLIKE ($\alpha = 0.10$; stationary block bootstrap with $l = \max(5, h)$).

Asset	h	fOU-BLUP	HAR-RV	ARFIMA	GARCH
000300	1	1.0233	1.0000*	1.0064*	1.2563
	5	1.0116*	1.0000*	1.0410*	1.0526*
	22	1.0010*	1.0000*	1.0068*	0.7792*
000016	1	1.0274*	1.0000*	1.0111*	1.2991
	5	1.0162*	1.0000*	1.0419*	1.1422*
	22	1.0117*	1.0000*	1.0107*	0.8567*
000905	1	0.9636*	1.0000	0.9780	1.6083
	5	0.9509*	1.0000	0.9868	1.5428
	22	0.9482*	1.0000	0.9591*	1.1665
399006	1	0.9722*	1.0000	0.9820	1.2248
	5	0.9398*	1.0000	0.9724	1.0465
	22	0.9387*	1.0000*	0.9707*	0.8317*
000001	1	1.0074*	1.0000*	1.0030*	1.4894
	5	0.9822*	1.0000*	1.0785*	1.1884
	22	0.9616*	1.0000*	0.9927*	0.8052*
QLIKE MCS90 rate	1	0.80	0.60	0.60	0.00
QLIKE MCS90 rate	5	1.00	0.60	0.60	0.40
QLIKE MCS90 rate	22	1.00	0.80	1.00	0.80

Online Appendix B. Cumulative losses, combinations, truncation, and pairwise DM

This appendix collects four low-cost robustness checks that leave the main-text MCS tables unchanged. First, Figure 10 plots cumulative squared forecast-error differentials

$$\text{CSFE}_t = \sum_{\tau=1}^t (e_{\tau, \text{HAR}}^2 - e_{\tau, \text{fOU}}^2) \quad (7.1)$$

for 510300 volume and CSI 300 RV at $h = 1$ and $h = 5$. An upward path means fOU-BLUP has lower cumulative squared error than HAR through date t . On 510300 volume the CSFE path rises steadily through the OOS window at both horizons, so the MSE ranking is not driven by a single crisis day. On CSI 300 RV the $h = 1$ path stays near zero with mixed sign—consistent with MCS co-membership and relative MSE near one against HAR-RV—while $h = 5$ accumulates a clear positive differential without a one-off jump.

$$\text{CSFE}_t = \sum_{\tau \leq t} (e_{\text{HAR}, \tau}^2 - e_{\text{fOU}, \tau}^2); \text{ upward favours fOU-BLUP}$$

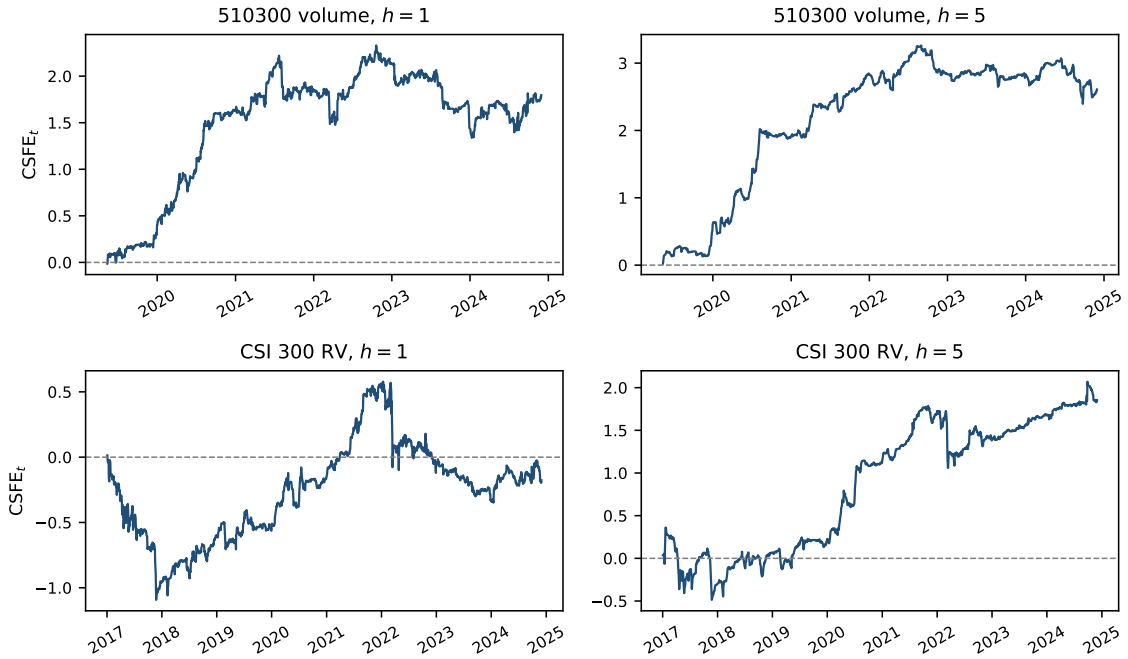


Figure 10: Cumulative squared forecast-error differential $\text{CSFE}_t = \sum_{\tau \leq t} (e_{\tau, \text{HAR}}^2 - e_{\tau, \text{fOU}}^2)$ for 510300 ETF volume (top) and CSI 300 index RV (bottom), at $h = 1$ and $h = 5$. Upward movement favours fOU-BLUP.

Second, Table XII reports an equal-weight combination $0.5 \text{fOU} + 0.5 \text{HAR}$ and a Mincer–

Zarnowitz encompassing regression of the realized target on the two forecasts (Mincer and Zarnowitz, 1969; Bates and Granger, 1969). On 510300 volume, fOU already dominates the equal-weight mix at $h = 1$, and $\hat{\beta}_1$ is significant while $\hat{\beta}_2$ is not, so HAR adds little incremental residual-scale information. On CSI 300 RV, the equal-weight mix slightly improves relative MSE and both slope coefficients are significant, consistent with complementary rough-spectral and HAR step-average information rather than nested predictors.

Table XII: Equal-weight combination and Mincer–Zarnowitz encompassing. Relative MSE is versus MA(22) for volume and versus HAR-RV for RV. The encompassing regression is $x_{t+h} = \alpha + \beta_1 \hat{x}_{t+h}^{\text{fOU}} + \beta_2 \hat{x}_{t+h}^{\text{HAR}} + \varepsilon_{t+h}$ with Newey–West t -statistics (lag $h - 1$).

Series	h	Rel. MSE			β_1 (fOU)		β_2 (HAR)	
		fOU	HAR	Equal	est.	t	est.	t
510300 volume	1	0.738	0.746	0.740	0.960	3.53	0.018	0.07
510300 volume	5	0.715	0.735	0.721	1.135	4.33	-0.194	-0.77
510300 volume	22	0.700	0.736	0.711	1.276	2.56	-0.379	-0.84
CSI 300 RV	1	1.001	1.000	0.994	0.490	3.04	0.509	3.25
CSI 300 RV	5	0.977	1.000	0.978	0.794	4.29	0.188	1.07
CSI 300 RV	22	0.962	1.000	0.966	0.833	2.95	0.069	0.25

Third, Table XIII varies the BLUP truncation length over $L \in \{20, 40, 60, 80\}$. Relative MSE stays within about one percent of the $L = 60$ benchmark on both representative series, so the main-text choice is not knife-edge.

Table XIII: Robustness of fOU-BLUP to truncation length L . Entries are relative MSE (volume versus MA(22); RV versus HAR-RV). The main-text predictor uses $L = 60$.

Series	h	$L = 20$	$L = 40$	$L = 60$	$L = 80$
510300 volume	1	0.739	0.738	0.738	0.738
510300 volume	5	0.719	0.714	0.715	0.715
510300 volume	22	0.705	0.701	0.700	0.701
CSI 300 RV	1	0.999	1.001	1.001	1.001
CSI 300 RV	5	0.972	0.976	0.977	0.975
CSI 300 RV	22	0.956	0.963	0.962	0.954

Fourth, Table XIV reports pairwise Diebold–Mariano statistics versus fOU-BLUP on squared-error loss, with the Harvey–Leybourne–Newbold small-sample factor (Harvey et al., 1997). Negative statistics favour fOU-BLUP. These pairwise tests complement the main-text MCS and do not replace it.

Table XIV: Pairwise Diebold–Mariano statistics versus fOU-BLUP on squared-error loss, with Harvey–Leybourne–Newbold (1997) small-sample adjustment. Negative values favour fOU-BLUP. Superscripts mark two-sided significance at 10% (*), 5% (**), and 1% (***). The main text leads with MCS; these pairwise tests are reported for completeness.

<i>Panel A. ETF volume</i>						
Asset	h	ARFIMA	HAR	AR(1)	ARMA	MA(22)
510300.SH	1	0.02	−1.69*	−7.00***	−4.34***	−9.24***
	5	−0.39	−2.44**	−8.73***	−4.61***	−5.10***
	22	−0.04	−1.45	−6.11***	−3.10***	−3.10***
510050.SH	1	0.58	−2.45**	−10.19***	−6.20***	−9.30***
	5	0.66	−1.96*	−13.09***	−6.22***	−4.95***
	22	0.39	−0.75	−10.68***	−4.84***	−2.77***
510500.SH	1	1.37	−1.41	−5.64***	−3.45***	−7.39***
	5	1.35	−1.20	−10.04***	−5.21***	−3.68***
	22	1.87*	−1.24	−9.38***	−5.02***	−2.65***
159915.SZ	1	−1.95*	−0.47	−6.87***	−5.08***	−9.32***
	5	−2.37**	−0.00	−9.08***	−7.75***	−5.14***
	22	−2.40**	0.42	−4.86***	−5.28***	−3.28***
159919.SZ	1	1.09	−0.62	−7.63***	−6.06***	−7.13***
	5	1.20	0.28	−11.27***	−9.65***	−3.35***
	22	1.91*	0.62	−6.15***	−5.38***	−0.97

<i>Panel B. Index RV</i>				
Asset	h	HAR-RV	ARFIMA	GARCH
CSI 300	1	0.19	0.73	−12.71***
	5	−1.59	−0.66	−7.39***
	22	−1.22	0.14	−3.80***
SSE 50	1	0.71	1.07	−13.96***
	5	−0.68	0.05	−8.25***
	22	−0.31	1.08	−4.28***
CSI 500	1	−2.42**	−0.98	−17.83***
	5	−1.88*	−0.24	−10.09***
	22	−1.80*	0.27	−5.31***
ChiNext	1	−3.46***	−0.88	−10.20***
	5	−4.29***	−1.25	−5.48***
	22	−3.61***	−1.00	−2.92***
SSE Comp.	1	−0.34	−0.01	−18.65***
	5	−2.28**	−1.42	−10.32***
	22	−1.97**	−0.63	−5.04***