

Volatility Persistence Is Not Distributional Memory: Separating Variance Memory from Empirical-Distribution Memory

Fuli Yang
GARCH Institute

Harvard Square, Cambridge, MA 02138, USA
ORCID: 0009-0005-4055-7561

Abstract

Volatility persistence is commonly summarized by the persistence parameter of a conditional variance recursion or by the autocorrelation of realized volatility. Neither statistic directly answers whether an empirical return distribution remains aligned with the same distributional state. This paper separates variance memory from distributional memory using daily returns on the S&P 500, FTSE 100, Nikkei 225, and Hang Seng from 2000 through June 2026. Distributional states are defined from pre-2007 low- and high-volatility prototypes, and are studied through state duration, Wasserstein displacement, and tail-state recurrence. The analysis combines a four-specification GARCH-family grid with a nested circular-block bootstrap that re-estimates prototypes and thresholds in each replication. All sixteen volatility-model fits converge, and Student-t GJR-GARCH has the lowest BIC in every market. The state-switching pattern is positive across the first three markets under the primary nested bootstrap, but only the Nikkei 225 retains a positive lower bound under the conservative 264-day block. Wasserstein and realized-volatility half-life comparisons are formally inconclusive, while no tail-recurrence lift survives Holm correction. The evidence supports a distinction between the estimands and a model-robust but block-sensitive pattern of state turnover; it does not establish a universal or causal separation between variance persistence and distributional memory. The numerical thresholds and decision rules were not changed in response to this revision.

Keywords: volatility persistence; distributional memory; GARCH; GJR-GARCH; Wasserstein distance; regime duration; structural breaks; bootstrap inference

JEL Classification: C22; C32; C52; G12

1. Introduction

A high estimate of GARCH persistence is often read as evidence that the market remains in the same volatility regime for a long time. That interpretation is understandable, but it combines two different empirical objects. A conditional variance recursion describes how a scale parameter updates. An empirical return distribution contains scale, location, asymmetry, tail thickness, and other features that can move without a proportional change in the conditional variance. The distinction matters whenever researchers use the word persistence as if it named a single property of market data.

This paper asks a deliberately narrow question: does high variance persistence imply distributional memory? The answer is not sought through another forecasting competition. Instead, I compare several clocks on a common 22-trading-day horizon. The first clock is the persistence of a conditional variance model. The second is realized-volatility autocorrelation. The third is the duration of a nearest-prototype distributional state. The fourth is the half-life of Wasserstein displacement from a low-volatility prototype. A fifth object, tail-state recurrence, is treated as a prespecified test rather than an affirmative result.

The empirical design is prospective and deliberately bounded. Prototypes are frozen using windows ending by 31 December 2006; evaluation begins in January 2007. The primary model is constant-mean Student-t GARCH(1,1), but the confirmatory audit also estimates Gaussian GARCH, Student-t GJR-GARCH, and variance-targeted Student-t GARCH. Inference uses circular moving blocks. The nested layer resamples both calibration and evaluation sequences and re-estimates the distributional prototypes and tail thresholds, so that uncertainty in the reference distributions is not treated as known.

Three findings organize the paper. First, the state-switching pattern is not confined to one symmetric Student-t specification: across the specification grid, the first three markets exhibit positive conservative lower bounds, while the Hang Seng does not. Second, allowing the calibration objects to vary makes the inference materially less precise. At the primary 132-day block, three markets pass the frozen screen; at 264 days, only the Nikkei 225 does. Third, the evidence does not establish that Wasserstein and realized-volatility decay rates are equivalent or

different under the prespecified equivalence band, and tail recurrence provides no surviving headline effect. These results are informative precisely because they limit the claim.

The paper contributes an estimand distinction. GARCH persistence is not invalid; it is a measure of variance memory. The claim is that it is incomplete as a measure of distributional state memory. The remainder proceeds as follows. Section 2 places the distinction in the literature. Section 3 defines the estimands. Section 4 describes the data and inference. Section 5 reports the specification grid and nested bootstrap. Section 6 discusses interpretation and limitations. Section 7 concludes.

2. Related literature

2.1 Conditional variance and realized volatility

Engle (1982) and Bollerslev (1986) established the conditional heteroskedasticity framework in which persistence is encoded in the propagation of past squared innovations and past conditional variance. In the standard GARCH(1,1), the sum of the ARCH and GARCH coefficients is a scale-memory parameter. Extensions such as the GJR-GARCH model allow negative innovations to affect conditional variance differently from positive innovations (Glosten et al., 1993). Realized-volatility research instead works with an ex post volatility measure constructed from high-frequency or daily returns, and studies its dependence through autocorrelation, long-memory specifications, and forecast comparisons (Barndorff-Nielsen and Shephard, 2002; Andersen et al., 2003; Hansen and Lunde, 2005; Corsi, 2009). Both traditions provide useful volatility clocks. Neither, by itself, defines persistence of the full return distribution.

2.2 Regime switching and structural instability

Hamilton (1989) formalized latent-state modeling for macroeconomic time series, and subsequent volatility research has used Markov-switching and related state-dependent specifications. A separate literature shows why persistence estimates can be sensitive to structural instability. Lamoureux and Lastrapes (1990) demonstrate that variance breaks can inflate apparent GARCH persistence, while Inclan and Tiao (1994) and Bai and Perron (2003) provide tools for detecting and representing changes in variance or regression parameters. This paper treats structural breaks as a competing qualification to the variance clock. It reports a BIC segmentation diagnostic, but does not claim that the diagnostic is a substitute for a complete formal break-test sequence.

2.3 Distribution-valued time series and Wasserstein methods

The distribution-valued time-series literature studies observations that are themselves probability distributions rather than scalar outcomes. Fréchet regression provides a general framework for regression with random objects (Agueh and Carlier, 2011; Petersen and Müller, 2019; Kokoszka et al., 2019), while Wasserstein autoregressive models provide a dynamic formulation for density-valued observations (Zhang et al., 2022; Panaretos and Zemel, 2020). Panaretos and Zemel (2020) discuss the statistical geometry of Wasserstein space. The present paper uses Wasserstein distance more modestly: as a transparent, one-dimensional distance between empirical windows and frozen prototypes. The objective is measurement and comparison, not a claim that the empirical windows follow a fully specified distribution-valued stochastic process.

2.4 Position of the contribution

The contribution is therefore not that variance models should be replaced by Wasserstein distances. It is that the two objects should not be interpreted as interchangeable. A variance recursion can remain persistent while the center, shape, or full-distribution displacement of rolling return windows changes. The empirical design makes this distinction testable by fixing the reference distributions before the evaluation period and by reporting uncertainty under alternative dependence blocks.

3. Estimands and hypotheses

3.1 Frozen distributional states

Let $F_{i,t}$ denote the empirical distribution of the 22 daily log returns in market i and window t . Within the calibration sample, I sort matured windows by realized volatility and form low- and high-volatility prototypes $Q_{i,L}$ and $Q_{i,H}$ as coordinatewise medians of sorted empirical quantile functions in the bottom and top

volatility quintiles. The prototypes are frozen after the calibration cutoff. For an evaluation window, define $d_{\{i,t\}}^{\{L\}} = W_1(F_{\{i,t\}}, Q_{\{i,L\}})$ and $d_{\{i,t\}}^{\{H\}} = W_1(F_{\{i,t\}}, Q_{\{i,H\}})$, where W_1 is the one-dimensional Wasserstein distance. The alignment score is $S_{\{i,t\}} = d_{\{i,t\}}^{\{L\}} / (d_{\{i,t\}}^{\{L\}} + d_{\{i,t\}}^{\{H\}})$; it increases as the window moves toward the high prototype. When both distances are close to zero, the window is close to both prototypes and the ratio can be numerically unstable; such near-coincidence is treated as an indeterminate boundary rather than as strong evidence for either state. Away from that boundary the denominator is positive and the score has the intended nearest-prototype interpretation. The high state is $S_{\{i,t\}} \geq 0.5$. This is a nearest-prototype classification, not a posterior probability.

3.2 Five clocks

The variance clock is the persistence of the conditional variance recursion. For a GARCH model it is $\alpha + \beta$; for GJR-GARCH it is $\alpha + \beta + \gamma/2$ under the usual symmetric-sign approximation. The realized-volatility clock is the lag-22 autocorrelation of $\log RV_{22}$. The state-duration clock is the median duration of high/low runs in phase-specific, non-overlapping sequences. The Wasserstein clock is the lag-22 autocorrelation of $D_{\{i,t\}} = W_1(F_{\{i,t\}}, Q_{\{i,L\}})$, converted to a half-life $h_D = 22 \log(0.5) / \log(\rho_D)$ when $0 < \rho_D < 1$. The tail clock measures recurrence of a new lower-tail state within three non-overlapping 22-day blocks after an exit, relative to a phase-matched unconditional benchmark.

3.3 Decision rules

The central comparison is not a test that correlation must be zero. The prespecified H1 gate asks whether at least two markets exhibit a scale-versus-shape persistence difference of at least 0.02; if triggered, segment-level estimates are used as the primary comparator for all four markets. H2 asks whether state switching remains detectable while the variance model remains persistent, with a positive lower confidence bound as the operational screen. H3 uses the equivalence band $[0.80, 1.25]$ for the ratio of Wasserstein to realized-volatility half-lives. H4 tests whether tail-state recurrence exceeds its phase-matched benchmark, using Holm correction across the 12 market-by-threshold tests. These rules are prospective from v0.2 onward; the project is not externally preregistered.

4. Data and empirical design

4.1 Markets and calendar treatment

The sample contains the S&P 500, FTSE 100, Nikkei 225, and Hang Seng price indices. Daily close-to-close percentage returns are $100 \times \log(P_t / P_{t-1})$ and cover January 2000 through June 2026. The series are Wind index-price series SPX.GI, FTSE.GI, N225.GI, and HSI.HI, downloaded on 5 August 2026. These are price indices rather than total-return indices. Dividends therefore enter as omitted downward adjustments to price returns. At the 22-day horizon this omission primarily changes the center and scale of each window by a relatively smooth market-specific component; it can still affect Wasserstein distances because W_1 is sensitive to location shifts. The analysis consequently interprets distances as distances between price-return distributions, not investor total-return distributions. A total-return replication is a useful future robustness exercise, but it is not silently substituted into the frozen sample. Each market is processed on its own observed trading calendar. Local holidays are not imputed, and the primary within-market estimands do not require cross-market date alignment.

4.2 Overlap-safe state duration

Adjacent daily rolling windows share 21 of 22 observations, so daily runs mechanically exaggerate duration. I therefore partition the evaluation sequence into 22 phase-specific sequences and sample windows every 22 trading days within a phase. Median runs are converted into trading days. The phase split and the bootstrap solve different problems. Phase-specific sampling removes the deterministic overlap created by adjacent rolling windows from the point estimate of duration; the circular moving-block bootstrap then supplies dependence-robust uncertainty for the resulting statistic. The 22 phases are therefore computational strata, not independent observations, and no phase-level independence assumption enters inference. The resulting duration statistic is a state-duration estimate at the common horizon, not a claim that a latent state changes only every 22 days.

4.3 Variance models and bootstrap

The baseline is a constant-mean standardized Student-t GARCH(1,1). The confirmatory grid adds Gaussian GARCH(1,1), standardized Student-t GJR-GARCH(1,1), and variance-targeted standardized Student-t

GARCH(1,1). Parameters are estimated separately by market. The evaluation-path switch rate is recomputed under each fitted conditional variance path. Following the dependent-bootstrap block-length literature (Politis and White, 2004; Patton et al., 2009), the primary bootstrap uses 1,999 circular moving-block replications with a 132-trading-day block. Sensitivity blocks are 66 and 264 trading days with 499 replications. The 264-day exercise is intentionally conservative but less precise: with 499 draws, a percentile endpoint is determined by only about the 13th and 487th ordered draws, compared with about the 50th and 1,950th draws under 1,999 replications. Wider intervals at 264 days therefore reflect both longer dependence blocks and lower Monte Carlo resolution; they should not be read as a pure test of market heterogeneity. In the nested layer, calibration-window and evaluation sequences are resampled independently; prototypes and tail thresholds are re-estimated in each replication. The random seed is 20260903.

4.4 Structural-break qualification

Before interpreting full-sample persistence, I run a Bai-Perron-style BIC segmentation diagnostic on log squared centered returns. Candidate break locations lie on a 22-day grid, the minimum segment length is 252 days, at most six breaks are allowed, and the penalty is twice BIC. The output is explicitly diagnostic rather than a complete Bai-Perron testing sequence. It is used to show how much the full-sample persistence comparison may depend on unmodeled variance instability.

5. Results

5.1 Structural instability qualifies the variance clock

The BIC segmentation diagnostic identifies five variance-location breaks in the S&P 500, FTSE 100, and Nikkei 225 evaluation samples and four in the Hang Seng. The median segment-level $\alpha + \beta$ estimates are 0.984, 0.902, 0.911, and 0.977, respectively. The reduction is large for the FTSE 100 and Nikkei 225 and more modest for the S&P 500 and Hang Seng. This result does not invalidate GARCH persistence; it shows that persistence can partly summarize a sequence of variance shifts. Because the family-level H1 gate is triggered, segment-level persistence is retained as the primary comparator, while full-sample values remain diagnostic.

5.2 GARCH-family specification grid

Market	Converged	Lowest-BIC model	Persistence	Switch-rate range
S&P 500	4/4	gjr_garch_t	0.986	4.8–8.1%
FTSE 100	4/4	gjr_garch_t	0.971	11.2–15.6%
Nikkei 225	4/4	gjr_garch_t	0.966	7.9–9.7%
Hang Seng	4/4	gjr_garch_t	0.987	2.4–4.0%

All sixteen fits converge. Student-t GJR-GARCH has the lowest BIC in every market. Switch rates range from 4.8% to 8.1% for the S&P 500, 11.2% to 15.6% for the FTSE 100, 7.9% to 9.7% for the Nikkei 225, and 2.4% to 4.0% for the Hang Seng. Under the conditional-on-calibration 264-day bootstrap, the lower confidence bound is positive for every specification in the first three markets and for none in the Hang Seng. The observed state-turnover pattern is therefore not produced solely by the symmetric Student-t baseline, although its inferential strength remains market- and block-dependent.

5.3 Nested calibration uncertainty

Market	132-day estimate	132-day 95% CI	264-day 95% CI
S&P 500	4.8%	[0.2%, 32.6%]	[0.0%, 35.0%]
FTSE 100	11.2%	[0.2%, 37.0%]	[0.0%, 36.6%]
Nikkei 225	7.9%	[0.7%, 34.8%]	[0.6%, 41.5%]
Hang Seng	4.0%	[0.0%, 21.8%]	[0.0%, 23.6%]

The nested bootstrap preserves the primary pattern: positive lower bounds for the S&P 500, FTSE 100, and Nikkei 225 at 132 days, but not for the Hang Seng. The conservative 264-day block widens the intervals substantially. Only the Nikkei 225 retains a positive lower bound. The appropriate interpretation is model-robust but block-sensitive evidence of turnover, not a universal separation between variance persistence and distributional state continuation.

5.4 Decay-rate comparison

The ratio of Wasserstein to realized-volatility half-lives is not statistically resolved by the nested bootstrap. None of the four 90% intervals lies wholly inside the equivalence band $[0.80, 1.25]$, so practical equivalence is not established. None lies wholly outside the band, so practical equivalence is not rejected. Point estimates close to one are therefore descriptive rather than confirmatory evidence.

5.5 Tail recurrence

Tail recurrence does not survive the prespecified inferential procedure. At the 85th, 90th, and 95th calibration thresholds, all nested confidence intervals include zero. After Holm correction across the 12 tests, no adjusted p-value rejects at 5%. The result is retained because it disciplines the narrative: the paper does not convert an isolated market-threshold pattern into a general claim about repeated tail states. The Hang Seng may differ because its constituent composition, trading-hours structure, and stronger links with mainland-China risk can produce a different distributional response to global shocks. These are interpretations, not identified mechanisms, and the present design cannot distinguish them.

6. Discussion

The central implication is interpretive. A high $\alpha + \beta$ or a high realized-volatility autocorrelation can be entirely meaningful while remaining silent about the persistence of the full return distribution. A market can preserve a slowly moving scale component while its location, asymmetry, or tail geometry shifts. Conversely, a distributional state can remain close to a prototype even when a scalar variance measure fluctuates within a broad scale band. The estimands answer related but non-identical questions.

The results also show why uncertainty about the calibration objects matters. If the low and high prototypes are treated as fixed, the state-switching pattern appears more precise. Once the calibration sequence is resampled and the prototypes and thresholds are re-estimated, the confidence intervals widen. This is not a nuisance detail: in a distributional-state design, the reference distributions are part of the measurement procedure. Treating them as known can overstate evidence for state turnover or persistence.

The structural-break diagnostic provides a second qualification. Very high full-sample persistence, especially near the IGARCH boundary, may absorb several variance mechanisms into one coefficient. The segment estimates reduce this concern for the FTSE 100 and Nikkei 225, but the diagnostic is not a complete formal break test. The paper therefore treats the reduction as a warning about interpretation, not as proof that all persistence is spurious.

Several limitations remain. The sample covers four equity indices, the primary window is 22 trading days, and the distributional state is a two-prototype construction. The nested bootstrap does not re-estimate GARCH parameters in every replication, and the 264-day sensitivity uses fewer replications. The phase-specific duration statistic is designed to reduce rolling-window overlap, but its discrete 22-day units create endpoint effects in confidence intervals. Finally, no result here is a causal regime estimate, a return-prediction result, or an investor-belief measure.

7. Conclusion

This paper separates a question often compressed into one word. GARCH persistence and realized-volatility autocorrelation measure memory in scale. Distributional state duration and Wasserstein displacement measure whether the empirical return law remains aligned with a reference state. In the four-market sample, a specification grid supports a model-robust pattern of state turnover in three markets, but nested inference shows that the strength of the result is sensitive to the dependence block and weaker in the Hang Seng. Half-life equivalence remains unresolved and tail recurrence is not supported. The conclusion is consequently modest but substantive: variance persistence is informative, yet it is not a complete description of distributional memory. The thresholds and decision rules used for this conclusion were fixed before the present revision and were not altered in response to the review.

Data and Materials Availability

The study uses Wind index-price series and computational procedures described in the manuscript. The retained analysis scripts, configuration file, and result tables are available from the author upon reasonable request, subject to data-licensing restrictions.

Ethics Approval and Consent to Participate

This study uses publicly available index-price data and involves no human participants, animals, personal data, or interventions. Ethics approval and informed consent were therefore not required.

Funder Statement

This research received no external funding.

Declaration of Interest

The author declares no financial, professional, or personal relationships that could have influenced the research, analysis, or interpretation reported in this manuscript.

Use of Generative AI

Generative AI tools were used for limited language editing, formatting assistance, and preliminary checks of document consistency. They were not used to generate data, estimate models, select results, determine statistical thresholds, or make substantive interpretations. The author reviewed and takes full responsibility for the manuscript, data construction, computational analysis, references, and conclusions.

References

- Agueh, M., & Carlier, G. (2011). Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2), 904–924. <https://doi.org/10.1137/100805741>
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2), 579–625. <https://doi.org/10.1111/1468-0262.00418>
- Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1), 1–22. <https://doi.org/10.1002/jae.659>
- Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, 64(2), 253–280. <https://doi.org/10.1111/1467-9868.00336>
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174–196. <https://doi.org/10.1093/jjfinec/nbp001>
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007. <https://doi.org/10.2307/1912773>
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5), 1779–1801. <https://doi.org/10.1111/j.1540-6261.1993.tb05128.x>
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357–384. <https://doi.org/10.2307/1912559>
- Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7), 873–889. <https://doi.org/10.1002/jae.800>
- Inclan, C., & Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89(427), 913–923. <https://doi.org/10.1080/01621459.1994.10476824>

- Kokoszka, P., Miao, H., Petersen, A., & Shang, H. L. (2019). Forecasting of density functions with an application to cross-sectional and intraday returns. *International Journal of Forecasting*, 35(4), 1304–1317. <https://doi.org/10.1016/j.ijforecast.2019.05.007>
- Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics*, 8(2), 225–234. <https://doi.org/10.1080/07350015.1990.10509794>
- Panaretos, V. M., & Zemel, Y. (2020). *An Invitation to Statistics in Wasserstein Space*. Springer.
- Patton, A., Politis, D. N., & White, H. (2009). Correction to “Automatic block-length selection for the dependent bootstrap” by D. Politis and H. White. *Econometric Reviews*, 28(4), 372–375. <https://doi.org/10.1080/07474930802459016>
- Petersen, A., & Müller, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Annals of Statistics*, 47(2), 691–719. <https://doi.org/10.1214/17-AOS1624>
- Politis, D. N., & White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1), 53–70. <https://doi.org/10.1081/ETC-120028836>
- Yang, F. (2026). Distributional regime persistence: Geometric market memory and its forecasting limits. SSRN Working Paper 7230659. <https://ssrn.com/abstract=7230659>
- Zhang, C., Kokoszka, P., & Petersen, A. (2022). Wasserstein autoregressive models for density time series. *Journal of Time Series Analysis*, 43(1), 30–52. <https://doi.org/10.1111/jtsa.12590>

Appendix A. Version chronology and decision-rule provenance

The project is not externally preregistered. The following chronology is supplied to make the boundary between design decisions and later diagnostics auditable. Version 0.1 froze the calibration cutoff, evaluation period, prototypes, primary horizon, and core estimands. Version 0.2 added numerical decision thresholds and the first break and bootstrap diagnostics in the same revision cycle; those thresholds are not claimed to predate those diagnostics. Version 0.3 clarified the H1 gate, endpoint behavior, and block-length scope. Version 0.4 completed the GARCH-family grid and nested calibration bootstrap. Version 0.5 converted the design package into a full manuscript. Version 0.6 responds to review by clarifying conditional inference, Monte Carlo precision, phase construction, price-index interpretation, and market-specific discussion. No threshold was changed in response to the v0.6 review.

Appendix B. Bootstrap interval convention and numerical precision

All reported bootstrap intervals are equal-tailed percentile intervals: the 2.5th and 97.5th empirical percentiles for 95% intervals and the 5th and 95th percentiles for 90% intervals. They are not BCa intervals. The 132-day primary results use 1,999 replications; the 66- and 264-day sensitivities use 499 replications. The smaller sensitivity run is sufficient for a directional robustness check but produces coarser percentile endpoints. In particular, a lower endpoint equal to zero can reflect both a broad sampling distribution and the finite resolution of the resampling experiment. This is why the 264-day results are reported as conservative sensitivity evidence rather than as a definitive ranking of markets.

Appendix C. H1-gate sensitivity and remaining scope

The H1 gate is a family-level routing rule: if at least two markets meet the prespecified scale-versus-shape difference criterion, segment estimates become the primary comparator for all four markets. This preserves a common cross-market estimand but makes the subsequent comparison conditional on the gate. The manuscript reports full-sample estimates as diagnostics and does not describe the gated intervals as unconditional coverage. A fully unconditional treatment would require a pre-specified selective-inference adjustment or a new design; neither is claimed here.

Appendix D. Data construction boundary

The data are daily price-index returns from Wind rather than total-return returns. This choice is held fixed across markets and across the calibration and evaluation samples. Because omitted dividends are not random at the annual horizon, the results should not be generalized to total investor wealth without a replication using total-return indices. The short common horizon limits, but does not eliminate, the resulting location component in Wasserstein comparisons.