

Forecast-Timing Conventions and the Value of Overnight Information in Volatility Forecasting*

Yi-Hao Lai[†]

July 2026

Abstract

Session-level volatility models produce overnight and intraday forecast components that are naturally issued at different times: the overnight component at the previous close, the intraday component at the market open, after the overnight return has realized. We show that the forecast-timing convention—when each component is issued and what the benchmark observes—drives the apparent value of session-level modeling. Using a parsimonious Periodic Realized GARCH (PRG, 6–8 parameters) on pinned data for six equity, commodity, and futures markets, we evaluate the same model under three conventions. Under a mixed-timing design that session-level evaluation easily slips into—model components issued at two different times, benchmark held at the previous close—PRG beats GJR-GARCH with Diebold–Mariano statistics of +4.3 to +6.4 in all six markets. Under a coherent close-time convention, with every component issued at the previous close, the advantage vanishes: zero of six markets clear the conservative $|t| > 3$ threshold, and the only nominally significant market points against PRG. Under a coherent open-time convention, where an information-matched GJR-X benchmark also observes the realized overnight return, PRG’s advantage is positive in all six markets and on the pinned vintage exceeds the conservative $|t| > 3$ threshold in five; the exception has the smallest overnight variance share. The same model and data thus deliver DM statistics from -2.3 to $+10.1$ depending only on the evaluation clock. Overnight information has genuine forecasting value at the open horizon and none at the day-ahead horizon; mixed-timing comparisons attribute to the model what belongs to the clock.

Keywords: Volatility forecasting; Forecast timing; Overnight information; Periodic GARCH; Forecast evaluation

*Computational infrastructure was provided by the author’s research group. All results derive from pinned snapshot data distributed with the replication package: daily series for six equity, commodity, and futures markets, each fixed to a single 2026-07-12 vintage and identified by SHA256 in the accompanying results JSON. Nothing is fetched live. The package reproduces every reported number from these snapshots and will be deposited in a public repository upon acceptance.

[†]Department of Finance, Dayeh University, Changhua, Taiwan. Email: yhlai@mail.dyu.edu.tw

JEL Classification: C22, C52, C53, G17

1 Introduction

Financial markets alternate between an overnight period, when exchanges are closed yet information arrives continuously and measurably moves next-session volatility (Tsiakas, 2008), and an intraday period of active price discovery whose volatility displays pronounced within-session periodicity (Andersen and Bollerslev, 1997). Blanc et al. (2014) document that the two sessions exhibit fundamentally different volatility feedback structures, and a substantial literature exploits session-level information for forecasting: periodic GARCH structures (Bollerslev and Ghysels, 1996), coupled intraday–overnight score-driven systems (Linton and Wu, 2020), continuous-time overnight diffusions (Kim et al., 2023), realized-variance models augmented with overnight returns (Todorova and Souček, 2014; Opschoor and Lucas, 2021), and regime-switching session models (Lai et al., 2024). These studies generally report that exploiting the session split improves volatility forecasts.

This paper examines a degree of freedom in such comparisons that has received little systematic attention: *when* each forecast component is issued. A session-level model naturally produces its overnight forecast at the previous close and its intraday forecast at the market open—after the overnight return has realized. Its full-day forecast is therefore a composite of components measurable with respect to two different information sets. Benchmarks such as GJR-GARCH, by contrast, issue a single day-ahead forecast at the previous close. Comparing the two-time composite against the one-time benchmark mixes a modeling question (does session structure help?) with an information question (what does acting at the open buy?). We show the mixture is not innocuous: it can manufacture the appearance of decisive model superiority where, at matched information sets, there is none.

Our vehicle is the Periodic Realized GARCH (PRG), a deliberately parsimonious model that extends the periodic GARCH of Bollerslev and Ghysels (1996) from calendar to session frequency: a single variance recursion alternates between session-specific parameter sets (6–8 parameters), so each session’s variance carries into the next. On pinned data for six markets—SPY, QQQ, GLD, EEM, the Taiwan 0050 ETF, and TAIEX index futures—we evaluate the identical model under three forecast-timing conventions. Under the *mixed* convention (model components at two times, benchmark at the close), PRG appears to dominate GJR-GARCH, with Diebold–Mariano statistics of +4.3 to +6.4 across all six markets. Under a coherent *close-time* convention, all components issued at the previous close, the advantage disappears in all six markets. Under a coherent *open-time* convention, against an information-matched GJR-X benchmark that also observes the realized overnight return, PRG wins in all six markets—five above the conservative $|t| > 3$ threshold on the pinned vintage, with the direction stable across vintages.

The contribution is threefold. First, we quantify, across six markets on a single pinned data vintage, how far the forecast-timing convention alone moves headline test statistics—from -2.3 to $+10.1$. Second, we provide a coherent dual-convention evaluation template: a close-time comparison for day-ahead applications and an information-matched open-time comparison for real-time updating. Third, the substantive finding is an honest localization of where overnight information pays: session-level modeling adds statistically significant value when forecasts are updated at the open—subject to a functional-form caveat we state explicitly in Section 4—and adds nothing over a plain GJR-GARCH for day-ahead forecasting.

2 Model and forecast-timing conventions

2.1 Session decomposition and the PRG model

Each trading day d splits into an overnight session ($s = 0$) and an intraday session ($s = 1$). With P_d^c and P_d^o the close and open prices,

$$r_{d,0} = \log(P_d^o/P_{d-1}^c), \quad r_{d,1} = \log(P_d^c/P_d^o), \quad (1)$$

and the common evaluation target for all models is the full-day variance proxy $\sigma_{\text{full},d}^2 = r_{d,0}^2 + r_{d,1}^2$, a whole-day variance assembled from session components in the spirit of Hansen and Lunde (2005). Under QLIKE loss, model rankings are robust to noise in unbiased proxies (Patton, 2011; Hansen and Lunde, 2006). For TAIEX, tick-level 5-minute realized variances replace squared session returns.

The PRG operates a single variance recursion over sessions indexed $n = 1, \dots, 2D$ with session indicator s_n :

$$h_n = \omega_{s_n} + \alpha_{s_n} x_{n-1} + \gamma_{s_n} x_{n-1} \mathbb{I}(r_{n-1} < 0) + \beta_{s_n} h_{n-1}, \quad (2)$$

where $x_n = r_n^2$ (the Basic variant sets $\gamma_s = 0$; “Realized” denotes the observed session-frequency proxy entering the recursion, distinct from the daily Realized GARCH of Hansen et al., 2012). Parameters are estimated by Gaussian QML (Bollerslev and Wooldridge, 1992). With $\rho_s = \alpha_s + 0.5\gamma_s + \beta_s$, covariance stationarity requires only $\rho_0\rho_1 < 1$ (Bollerslev and Ghysels, 1996). The cross-session term $\beta_{s_n} h_{n-1}$ is the *session bridge*: each session’s conditional variance feeds the next.

2.2 Three forecast-timing conventions

Let $\mathcal{F}_{d-1}^c$ be the information set at the day- $(d-1)$ close and $\mathcal{F}_d^o$ the set at the day- d open; $\mathcal{F}_d^o$ contains the realized overnight return $r_{d,0}$. The PRG session forecasts are

$$\hat{h}_{d,0} = \mathbb{E}[r_{d,0}^2 \mid \mathcal{F}_{d-1}^c], \quad (3)$$

$$\hat{h}_{d,1}(z) = \omega_1 + \alpha_1 z + \gamma_1 z \mathbb{I}(r_{d,0} < 0) + \beta_1 \hat{h}_{d,0}, \quad (4)$$

where z is whatever stands in for the overnight squared return. Three full-day forecast objects follow:

- **Close convention (coherent).** Everything is issued at the day- $(d-1)$ close: $\hat{\sigma}_C^2 = \hat{h}_{d,0} + \hat{h}_{d,1}(\hat{h}_{d,0})$, with the unobserved overnight square replaced by its model-consistent expectation (leverage indicator at its symmetric expectation $\frac{1}{2}$). Every input is $\mathcal{F}_{d-1}^c$ -measurable. The natural benchmark is GJR-GARCH (Glosten et al., 1993) on close-to-close returns, also issued at $\mathcal{F}_{d-1}^c$.
- **Open convention (coherent).** The full-day object is formed at the day- d open: $\hat{\sigma}_O^2 = r_{d,0}^2 + \hat{h}_{d,1}(r_{d,0}^2)$, with the realized overnight component plugged in. An information-*matched* benchmark must also observe $r_{d,0}$: we use GJR-X, a daily GJR augmented with the *current-day* overnight squared return, $h_t = \omega + \alpha r_{t-1}^2 + \gamma \mathbb{I}(r_{t-1} < 0) r_{t-1}^2 + \beta h_{t-1} + \delta r_{\text{overnight},t}^2$, issued at the same open.
- **Mixed convention (incoherent).** The composite $\hat{\sigma}_M^2 = \hat{h}_{d,0} + \hat{h}_{d,1}(r_{d,0}^2)$ combines an $\mathcal{F}_{d-1}^c$ -measurable component with an $\mathcal{F}_d^o$ -measurable one, and is compared against a benchmark held at $\mathcal{F}_{d-1}^c$. The composite corresponds to no coherent single-issuance-time forecast—it treats $r_{d,0}$ as unobserved in its first component yet observed in its second, choices no single information set makes simultaneously rational—and it is benchmarked against a model held at the earlier set $\mathcal{F}_{d-1}^c$, which is handicapped by construction. This is the convention implicit in evaluations that sum session forecasts “as produced” and compare against day-ahead models—including earlier drafts of this paper.

The mixed convention is not a straw man: session-level models naturally produce their components at different issuance times, single-recursion benchmarks are issued once a day, and nothing in standard evaluation practice forces the two information sets to be matched—earlier drafts of this paper reported exactly this comparison. Our point is that the resulting comparison confounds model structure with information timing, and Section 4 quantifies the confound.

2.3 Evaluation

Losses are QLIKE; pairwise comparisons use Diebold and Mariano (1995) tests with Newey–West HAC variance (bandwidth $\lceil n^{1/3} \rceil$) and the Harvey et al. (1997) small-sample correction. With the 18 pairwise DM tests of Table 2 as the reported family ($\alpha/m = 0.05/18 \approx 0.0028$, two-sided $|z| \approx 2.99$) we adopt $|t| > 3.0$ as a conservative, Bonferroni-style significance threshold; the six lagged-variant robustness tests of Section 4 are reported as diagnostics outside this family, since they re-examine the same close-time comparison under an alternative plug-in rather than adding an independent hypothesis; we report nominal p -values throughout so readers can apply their own standard. All estimation is strictly out-of-sample with periodic refitting (GJR and GJR-X every 63 trading days; PRG every 126–252 sessions), using only data before each forecast origin.

3 Data

Table 1 summarizes the six markets. All OHLC series are dividend- and split-adjusted prices pinned as snapshot files (vintage 2026-07-12), so overnight returns exclude predictable ex-dividend jumps; every number in this paper is bound to the archived pinned-vintage result files, which the replication package reproduces from the snapshots.¹

Table 1: Data and session decomposition

Market	Source	OOS obs	OOS period	ON var share (%)
SPY	OHLC (pinned)	1,823	2019-01–2026-04	44.8
QQQ	OHLC (pinned)	1,981	2018-05–2026-04	38.5
GLD	OHLC (pinned)	1,613	2019-10–2026-04	60.9
EEM	OHLC (pinned)	1,734	2019-05–2026-04	70.7
0050.TW	OHLC (pinned)	1,251	2021-01–2026-04	63.5
TAIFEX TX	Tick (5-min RV)	843	2022-07–2025-12	68.9

Notes: OOS obs is the out-of-sample day count (70/30 split for OHLC markets). ON var share is the overnight share of full-day variance over the OOS window, computed from the pinned snapshots. TAIFEX uses tick-level 5-minute realized variance for the overnight (15:00–08:45) and regular (08:45–13:45) sessions of the volume-selected near-month contract; 0050.TW is cleaned for the 2014 split.

¹Earlier circulated drafts reported results from unpinned, vintage-drifting downloads (e.g., an SPY headline DM of 6.00 under the mixed convention). All results here derive from the pinned snapshots; the replication package documents the vintage, adjustment convention, and estimation seeds.

4 Results

4.1 The flip

Table 2 is the paper’s central result: the same PRG specification, on the same pinned data, evaluated under the three conventions of Section 2.2.

Table 2: The forecast-timing flip: DM statistics for the same model and data under three conventions

Market	N	Mixed PRG vs. GJR (incoherent)	Close PRG vs. GJR (both at $\mathcal{F}_{d-1}^c$)	Open PRG vs. GJR-X (both at $\mathcal{F}_d^o$)
SPY	1,823	+5.83*** (<0.001)	−0.74 (0.46)	+3.56*** (<0.001)
QQQ	1,981	+4.78*** (<0.001)	−2.28 (0.02)	+1.56 (0.12)
GLD	1,613	+6.11*** (<0.001)	+0.44 (0.66)	+3.64*** (<0.001)
EEM	1,734	+6.40*** (<0.001)	+0.54 (0.59)	+10.14*** (<0.001)
0050.TW	1,251	+5.19*** (<0.001)	+0.32 (0.75)	+3.67*** (<0.001)
TAIFEX	843	+4.33*** (<0.001)	+0.49 (0.62)	+5.50*** (<0.001)
Harvey-significant ($ t > 3$)		6/6	0/6	5/6

Notes: PRG Extended specification (Eq. 2) throughout. DM t -statistics on QLIKE loss differentials, reported as benchmark loss minus PRG loss, so positive values favor PRG; nominal two-sided p -values in parentheses; *** marks the conservative $|t| > 3.0$ threshold. Mixed: PRG composite $\hat{h}_{d,0} + \hat{h}_{d,1}(r_{d,0}^2)$ vs. GJR issued at the close. Close: PRG composite $\hat{h}_{d,0} + \hat{h}_{d,1}(\hat{h}_{d,0})$ vs. GJR, both issued at the day- $(d-1)$ close. Open: PRG composite $r_{d,0}^2 + \hat{h}_{d,1}(r_{d,0}^2)$ vs. GJR-X with the current-day overnight regressor, both issued at the day- d open. DM variances use Newey–West HAC with bandwidth $\lceil n^{1/3} \rceil$; doubling the bandwidth changes no verdict. All three panels use the identical pinned data vintage and deterministic estimation seeds.

Read across any row. Under the mixed convention, PRG appears to dominate GJR decisively in every market—exactly the kind of headline the overnight-information literature reports. Under the coherent close-time convention, the advantage evaporates: DM statistics straddle zero, no market comes near the conservative threshold, and the only nominally significant cell (QQQ, $t = -2.28$, $p = 0.02$) points *against* PRG. Under the coherent open-time convention—where the benchmark is granted the same realized overnight return—PRG’s advantage is positive in all six markets and exceeds $|t| > 3$ in five; the exception is QQQ ($t = +1.56$), the market with the smallest overnight variance share. The spread between conventions within a single market’s row reaches 9.6 t -units (EEM). Nothing about the model or the data changes across the panels; only the clock does.

4.2 Interpretation

What mixed-timing comparisons measure. The mixed panel’s inflation has a mechanical source: the intraday component $\hat{h}_{d,1}(r_{d,0}^2)$ absorbs the realized overnight shock while the GJR benchmark, stuck at the previous close, cannot. The gap between the mixed and close panels—3.8 to 7.1 t -units per market—

is the part of the headline that belongs to the information clock, not to session modeling. Studies that sum session-level forecasts “as issued” and benchmark them against day-ahead models inherit this confound, and their reported significance levels should be read accordingly.

Where overnight information genuinely pays. The open panel shows that the value of session modeling at the open horizon is real, not an artifact of information mismatch: GJR-X sees the same realized overnight return, yet PRG’s session-specific propagation of that information—through parameters estimated separately for the overnight and intraday processes and linked by the bridge—beats the exogenous-regressor route in all six markets. The cross-market pattern is itself informative: on the pinned vintage the open-time DM statistic is perfectly rank-ordered in the market’s overnight variance share, from EEM (70.7%, $t = +10.14$) down to QQQ (38.5%, $t = +1.56$, the only insignificant market)—descriptive with six markets, vintage-sensitive at the low-share end (see Robustness below), but directionally the ordering a genuine overnight-information channel predicts. One caveat keeps this honest: part of the gap may reflect functional form rather than information processing, since PRG embeds the realized overnight component additively in its composite while GJR-X must load it through a single linear regressor. An intraday-target-only design that removes the additive component from both sides is left for follow-up.

No exception in high-overnight-share markets. A natural conjecture is that markets whose overnight session carries most of the variance (EEM at 70.7%, TAIEX at 68.9%, 0050.TW at 63.5%, GLD at 60.9%) might retain a close-time PRG advantage. They do not: their close-time DM statistics are +0.54, +0.49, +0.32, and +0.44—directionally favorable, nowhere near significance.

Robustness. The close-time null is robust to how the unobserved overnight square is plugged in: a lagged-realized variant ($z = r_{d-1,0}^2$) leaves the verdict unchanged: no market clears the conservative threshold in either variant. The one cell that comes close, QQQ, strengthens from $t = -2.28$ to -2.95 ($p = 0.003$)—still short of the threshold and still against PRG—and the near-zero 0050.TW cell flips sign across variants (+0.32 vs. -0.28), exactly as statistical noise should. Doubling the HAC bandwidth changes no verdict in any panel. Point statistics—QLIKE levels and individual DM values—are sensitive to data vintage and to estimation basins of the non-convex likelihood: an earlier unpinned run of the open-time comparison put SPY at +2.1 and QQQ at +3.0 versus the pinned +3.56 and +1.56 here. No direction flips in any market under any vintage, and the pinned values reproduce bit-identically from the archived snapshots. Two headline framings are themselves vintage-sensitive and should be read as pinned-vintage statements: the open-panel Harvey count (five of six here; four of six in the pilot vintage,

where SPY sits below the threshold) and the perfect share-ordering (the pilot inverts the two lowest-share markets). The vintage-robust statements are the direction—positive in all six markets—and the large-margin, high-share markets (EEM, TAIEX); we make no claims from raw QLIKE rankings or borderline point values.

5 Discussion and conclusion

This paper’s finding is a methodological warning with a constructive core. The warning: forecast-timing conventions are not a technical footnote. The same session-level volatility model, on the same pinned data, delivers Diebold–Mariano statistics anywhere from -2.3 to $+10.1$ depending only on when forecast components are issued and what the benchmark observes. Comparisons that combine components issued at different times—or that benchmark open-informed forecasts against close-informed models—can overstate model value by several t -units per market. Reported significance under such designs conflates the model with the clock. The constructive core: evaluated coherently, overnight information has a well-defined and horizon-specific value. For day-ahead forecasting at the previous close, session-level structure adds nothing over a plain GJR-GARCH: the overnight return is not yet observed, and forecasting it session-by-session is no better than forecasting the day directly. For real-time updating at the open, session structure adds value that an exogenous-regressor patch does not replicate—significantly so in five of six markets on the pinned vintage, increasingly so the larger the market’s overnight variance share, and subject to the functional-form caveat of Section 4.

The practical prescription for researchers is to report the issuance time of every forecast component and match the benchmark’s information set to it—our close/open panel pair provides a template. The prescription for practitioners is equally direct: a session-level model such as PRG is a tool for the open, not for the close. PRG itself remains attractive as a vehicle—6–8 parameters, standard QML, no latent regimes (Lai et al., 2024; Haas et al., 2004)—precisely because its parsimony makes the convention experiment clean: nothing in the flip can be attributed to estimation complexity.

Limitations: OHLC markets use squared session returns as variance proxies (QLIKE rankings are proxy-robust, Patton, 2011); tail-risk and economic-value evaluation under the two coherent conventions, and an intraday-target-only comparison isolating functional form from information, are natural follow-ups.

References

- Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2–3), 115–158. [https://doi.org/10.1016/S0927-5398\(97\)00004-2](https://doi.org/10.1016/S0927-5398(97)00004-2)
- Blanc, P., Chicheportiche, R., and Bouchaud, J.-P. (2014). The fine structure of volatility feedback II: Overnight and intra-day effects. *Physica A*, 402, 58–75. <https://doi.org/10.1016/j.physa.2014.01.047>
- Bollerslev, T. and Ghysels, E. (1996). Periodic autoregressive conditional heteroskedasticity. *Journal of Business & Economic Statistics*, 14(2), 139–151. <https://doi.org/10.1080/07350015.1996.10524640>
- Bollerslev, T. and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2), 143–172. <https://doi.org/10.1080/07474939208800229>
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. <https://doi.org/10.1080/07350015.1995.10524599>
- Glosten, L. R., Jagannathan, R., and 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>
- Haas, M., Mittnik, S., and Paoletta, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4), 493–530. <https://doi.org/10.1093/jjfinec/nbh020>
- Hansen, P. R. and Lunde, A. (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics*, 3(4), 525–554. <https://doi.org/10.1093/jjfinec/nbi028>
- Hansen, P. R. and Lunde, A. (2006). Consistent ranking of volatility models. *Journal of Econometrics*, 131(1–2), 97–121. <https://doi.org/10.1016/j.jeconom.2005.01.005>
- Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6), 877–906. <https://doi.org/10.1002/jae.1234>

- Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291. [https://doi.org/10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4)
- Kim, D., Shin, M., and Wang, Y. (2023). Overnight GARCH-Itô volatility models. *Journal of Business & Economic Statistics*, 41(4), 1215–1227. <https://doi.org/10.1080/07350015.2022.2116027>
- Lai, Y.-H., Wang, Y.-C., and Chang, Y.-C. (2024). Forecasting trading-session return volatility in Taiwan futures market: A periodic regime switching with jump approach. *Asia-Pacific Financial Markets*, 31(2), 285–305. <https://doi.org/10.1007/s10690-023-09415-w>
- Linton, O. and Wu, J. (2020). A coupled component DCS-EGARCH model for intraday and overnight volatility. *Journal of Econometrics*, 217(1), 176–201. <https://doi.org/10.1016/j.jeconom.2019.12.015>
- Opschoor, A. and Lucas, A. (2021). Observation-driven models for realized variances and overnight returns applied to Value-at-Risk and Expected Shortfall forecasting. *International Journal of Forecasting*, 37(2), 622–633. <https://doi.org/10.1016/j.ijforecast.2020.07.009>
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246–256. <https://doi.org/10.1016/j.jeconom.2010.03.034>
- Todorova, N. and Souček, M. (2014). Overnight information flow and realized volatility forecasting. *Finance Research Letters*, 11(4), 420–428. <https://doi.org/10.1016/j.frl.2014.07.001>
- Tsiakas, I. (2008). Overnight information and stochastic volatility: A study of European and US stock exchanges. *Journal of Banking & Finance*, 32(2), 251–268. <https://doi.org/10.1016/j.jbankfin.2007.03.008>