Highlights

Do option-implied signals improve forecasts of realized correlation?

Tomáš Plíhal, Joachim Oliver Mampouya

- The correlation risk premium lifts HAR forecasts of realized correlation
- A history-free premium and a long-memory benchmark trace the gain to realized memory
- The premium adds nothing once the benchmark carries the year its implied leg embeds
- The premium repairs covariance-matrix coherence; a physical-beta placebo fails
- With costs charged, four standard uses pay no fee: the value is coherence, not accuracy

Do option-implied signals improve forecasts of realized correlation?

Tomáš Plíhal[a,*], Joachim Oliver Mampouya[a]

[a]Department of Finance, Faculty of Economics and Administration, Masaryk University, Lipová 41a, Brno, 602 00, Czech Republic

ARTICLE INFO

Keywords:
Realized correlation
Option-implied information
Correlation risk premium
HAR
Out-of-sample forecasting

ABSTRACT

Option-implied information improves out-of-sample forecasts of monthly realized correlation for S&P 500 equity pairs, but the improvement turns out not to be forward-looking. We enrich the standard HAR benchmark with option-implied feature groups, at their core the pairwise correlation risk premium: implied correlation, recovered from the two stocks' option-implied betas, less the realized correlation the benchmark already carries. The premium raises the out-of-sample R^2 by 4.22% over 14,144,670 pair-months, a significant gain. Tests of the source of that improvement, however, point to backward-looking history rather than option information: a version of the premium built from betas that use current option prices alone forecasts nothing, and a realized benchmark handed a year of the pair's own correlation history absorbs the premium entirely. The gain is long realized-correlation memory hiding inside the implied inputs, a caution for any implied-correlation measure built on historically anchored components. What option prices do supply is structural. Covariance matrices assembled from hundreds of thousands of pairwise forecasts are frequently not positive definite; adding the premium repairs that incoherence, where a placebo with equal forecasting content fails, and delivers the only significant reductions in realized minimum-variance portfolio risk in our model grid. With trading costs charged, the repair earns no fee across four standard applications: the value of option-implied correlation in portfolio choice lies in reconciling many forecasts, not in sharpening any one.

1. Introduction

The risk of a diversified portfolio depends more on the correlations among its holdings than on their individual volatilities, and those correlations move. They rise in downturns but not in upturns (Longin and Solnik, 2001; Ang and Chen, 2002), and they have trended upward over decades as tail dependence has risen with them (Christoffersen, Errunza, Jacobs and Langlois, 2012). Diversification is therefore weakest in the states where it is most needed. Investors are compensated for bearing that risk. Driessen, Maenhout and Vilkov (2009) show that individual variance risk is not priced once index variance risk is controlled for, and that what is priced is correlation; Krishnan, Petkova and Ritchken (2009) reach the same conclusion from the cross-section of stock returns, Buraschi, Kosowski and Trojani (2014) from hedge-fund returns, and Mueller, Stathopoulos and Vedolin (2017) from currencies. Expected correlation therefore has a price attached, and forecasting it is a question about a priced state variable.

The forecasting toolkit for that object is built almost entirely from past prices. High-frequency data deliver realized covariances and correlations with a distribution theory behind them (Barndorff-Nielsen and Shephard, 2004). The heterogeneous autoregressive (HAR) model of Corsi (2009) was designed for realized volatility, capturing its persistence through daily, weekly, and monthly lags, but the cascade structure carries over to correlations, and Audrino and Corsi (2010) apply it to realized correlations directly. Multivariate extensions address the joint structure and the measurement error in it (Chiriac and Voev, 2011; Bollerslev, Patton and Quaedvlieg, 2018). Decomposing comovement by the sign of the underlying returns adds further predictive content, since correlation behaves differently in joint downturns (Bollerslev, Li, Patton and Quaedvlieg, 2020; Patton and Sheppard, 2015; Bollerslev, Patton and Quaedvlieg, 2022). The most complete design of this kind is that of Bollerslev, Li and Tang (2026), who take a HAR in lagged correlations as the benchmark to beat and forecast monthly realized correlations for equity pairs from a large set of realized features, among them correlations obtained by projecting the realized covariance matrix onto observable firm characteristics in the manner of Fama and French (2020). Every one of these predictors is backward-looking.

Option prices, by contrast, are forward-looking. Model-free implied volatility subsumes both its Black–Scholes counterpart and past realized volatility as a forecast of future volatility (Jiang and Tian, 2005), and the difference

*Corresponding author
tomas.plihal@econ.muni.cz (T. Plíhal); joachim.oliver.mampouya@econ.muni.cz (J.O. Mampouya)
ORCID(s):

between risk-neutral and realized variance, a risk premium formed as implied minus realized (Carr and Wu, 2009), predicts returns in its own right (Bollerslev, Tauchen and Zhou, 2009). For correlation the measurement side is equally well developed. At the market level, implied correlation can be read from index and constituent options (Skintzi and Refenes, 2005; Driessen et al., 2009) or extracted model-free as a full dependence structure (Bondarenko and Bernard, 2024); for an individual pair of stocks, no traded option quotes it directly, so it is recovered from the two stocks' risk-neutral betas, the construction of Buss and Vilkov (2012) to which we return below. The wedge between index and single-name option prices that these measures read carries genuine economic information rather than a pricing artifact (Kelly, Lustig and Van Nieuwerburgh, 2016), option-implied information prices the equity cross-section directly (Goyal and Saretto, 2009; An, Ang, Bali and Cakici, 2014), and forecasting with it has a survey literature of its own (Christoffersen, Jacobs and Chang, 2013; Engle and Figlewski, 2015).

What has not been done is to bring the two sides together. Bollerslev et al. (2026) deliberately omit implied features from their design on option-coverage grounds and say so. In portfolio selection the record is split by channel: DeMiguel, Plyakha, Uppal and Vilkov (2013) find that option-implied volatilities and skewness improve portfolios while implied *correlations* do not, and Kempf, Korn and Saßning (2015) reach the same conclusion for minimum-variance portfolios built on implied correlations alone. But a portfolio optimization inverts a covariance matrix, and inversion is dominated by estimation error in the inputs (DeMiguel, Garlappi and Uppal, 2009; Bollerslev et al., 2018), so a null there is weak evidence about the forecast. This paper asks the forecasting question directly: does forward-looking, option-implied information improve out-of-sample forecasts of monthly realized pairwise correlation beyond what backward-looking realized measures already capture, and if so, does the improvement carry economic value?

In this paper we run that comparison on one sample, at scale, and under a single evaluation rule. We build monthly realized correlations for pairs of S&P 500 members from 15-minute intraday returns over 2005–2023, 770 stocks in all, and forecast them one month ahead with a grid of 26 models: four realized rungs of increasing strength, five option-implied feature groups augmenting each, and a penalized pair selecting over everything. Every model is scored on a single common sample of 14,144,670 pair-months over 168 response months, with inference on monthly aggregates of the loss differential rather than on millions of dependent pair-months.

On the implied side, our organizing principle is that option-implied signals enter as *premia*. Our core feature is the pairwise correlation risk premium, implied correlation less the realized correlation the benchmark already carries, together with its market-wide counterpart. We extract the pairwise implied correlation from the widely used option-implied betas of Buss and Vilkov (2012). These betas, however, are not built from option prices alone: the construction corrects a historical correlation matrix, estimated from roughly a year of past returns, toward its risk-neutral counterpart, an anchoring that Hollstein and Prokopczuk (2016) identify as why it wins beta horse races. The premium therefore arrives carrying an embedded year of realized correlation history, and its gain over a short-memory benchmark could be the option market's information or simply that memory.

We treat this as the paper's central identification problem and answer it twice. First, we rebuild the premium from the Martin and Wagner (2019) implied betas, which are functions of current option prices alone; this version carries only option information, so whatever it forecasts is forward-looking. Second, we control for the memory directly: our strongest realized benchmark adds trailing means of the pair's own realized correlation, so whatever an implied feature contributes on top of it cannot be that memory. Option information should survive both tests; inherited history should survive neither.

Augmenting the HAR benchmark with the correlation risk premium raises the out-of-sample R^2 by 4.22%, a statistically significant gain. Both tests, however, return the same answer. The history-free version forecasts nothing: built identically but without the embedded return history, it improves on HAR by 0.49%, statistically indistinguishable from zero. And the long-memory benchmark, built from realized history alone, beats HAR by 7.94%, nearly double the premium's gain, while on top of that benchmark the premium's marginal contribution falls to −0.76%; a penalized regression allowed to select from the entire implied feature block agrees. Identified from both directions, the premium's forecasting gain over HAR is long realized-correlation memory that the HAR family lacked, not forward-looking information.

Any option-implied correlation measure built on a historical input embeds that history, so when such a measure outperforms a realized benchmark with a shorter memory, part of the measured gain may be the embedded past rather than option information. Two further facts sharpen the reading. The gain is flat across the size distribution and significant among the largest pairs, so it is not a small-pair artifact. And the component that fades over the sample is the priced-correlation channel specifically: the premium's contribution declines across calendar blocks while the implied-beta group runs the other way.

The answer to the economic half is a boundary with one option-specific property inside it. A covariance matrix assembled from hundreds of thousands of pooled pairwise forecasts is frequently not positive definite, and that incoherence has a cost: across our grid, the models whose matrices fail most often are the models whose minimum-variance portfolios realize the most risk. Adding the correlation risk premium *repairs* the matrix, and the repair delivers the grid's only significant reductions in realized portfolio risk, on every premium-augmented rung once the market factor is constrained out of both sides of the comparison. A placebo pins the property to option prices: pair features built from sixty-month *physical* betas match the premium's forecasting content yet destroy coherence, and the history-free premium breaks the matrix outright. Option prices contribute structure rather than accuracy. They are what allows thousands of simultaneous pairwise forecasts to assemble into a usable covariance matrix.

It is a repair an investor is not paid for. With the ten-basis-point trading charge applied throughout (a charge that runs in the premium's favor, since it turns over less than its benchmark), the sharper forecast earns no fee on any of the four standard uses of a correlation forecast. Collapsed into an average predicted correlation, it does not sharpen equity-premium prediction, and the purely forward-looking market premium predicts nothing at all. Inverted into minimum-variance weights, it lowers realized risk but gives up more in mean return than the reduction is worth. Read as a cross-sectional ranking, it orders pairs-trading strategies as forecast accuracy predicts, and none of them earns an alpha distinguishable from zero. And asked to price a dispersion bet, it understates the risk by the widest margin, because the common structure that repairs the matrix also compresses the cross-sectional dispersion that a long-short spread trades on.

The paper contributes on three fronts. First, to the literature that extracts correlations from option prices (Driessen et al., 2009; Buss and Vilkov, 2012; Bondarenko and Bernard, 2024), it crosses the implied and realized-correlation forecasting toolkits on one sample, the crossing Bollerslev et al. (2026) left open, and shows that the measured "option-implied" forecasting gain decomposes, by construction and by control, into inherited realized memory. The decomposition reinterprets any implied-correlation result built on a historically anchored measure, and it was obtainable only because the long-memory benchmark was run. Second, to the literature on the economic value of correlation forecasts (DeMiguel et al., 2013; Kempf et al., 2015), it identifies by placebo the one margin where option prices contribute something no realized substitute in the design does, namely coherence across thousands of simultaneous forecasts, and shows that this structural value does not convert into a fee on any of four standard margins, with each null carrying its mechanism. Third, the measurement choices are what let the first two claims be read at face value: one common evaluation sample across all 26 models, inference on monthly aggregates, trading costs charged wherever a return is compared, and a studentized family-level test reported in its conservative form.

We are also explicit about scope. Our evaluation window begins after the rolling estimation window is consumed and so excludes the 2008 crisis, where out-of-sample equity-premium predictability is known to concentrate (Rapach, Strauss and Zhou, 2010), and Buss, Schönleber and Vilkov (2019) report that expected correlation extracted jointly from index and single-name options does predict market returns at longer horizons over a crisis-inclusive sample. Our aggregate null is a statement about a monthly horizon and a post-crisis sample, and we read it as one. Nor can 168 months separate a fading signal from a turbulent final block, or either from the memory channel; we report the decline without choosing among them.

Section 2 develops the realized, factor-driven, and option-implied features, the models, and the evaluation framework. Section 3 describes the sample and reports descriptive statistics. Section 4 presents the out-of-sample forecasting results and the in-sample coefficients behind them. Section 5 takes the forecasts to the four economic applications. Section 6 reports subsample stability, the size profile of the gain, and the strategy parameter sweeps. Section 7 concludes.

2. Methodology

This section builds the paper's inputs in order of increasing forward-lookingness: the realized correlation measures (Section 2.1), their projection onto observable firm characteristics (Section 2.2), and the option-implied features (Section 2.3), which the models of Section 2.4 combine in one common linear form. Section 2.5 gives the estimators and the rolling out-of-sample design, and Section 2.6 the criteria on which every model is scored. One discipline runs through all six parts: every feature is measured through the end of month t, every model shares the same functional form and estimation windows, and every forecast is judged on one common sample. The constructions specific to each economic application, the portfolio weights and trading rules, are presented in Section 5 beside the results they produce; this section carries only what every result shares.

2.1. Realized correlation measures

We construct realized variances, covariances, and correlations from intraday returns, following Bollerslev et al. (2026). Let $p_{i,d,k}$ denote the log price of stock i at the k-th intraday grid point on day d. Within each regular trading session (09:30–16:00 ET) we sample trade-prices on a 15-minute grid, yielding 27 points (the 09:30 open and the 26 fifteen-minute marks 09:45, 10:00, …, 16:00) and hence 26 intraday returns $r_{i,d,k} = p_{i,d,k} - p_{i,d,k-1}$, $k = 1, \dots, 26$. Following common practice (Hansen and Lunde, 2005), we append the overnight return $r^{\text{on}}_{i,d} = p_{i,d,0} - p_{i,d-1,26}$, today's open less the previous session's close, for a full day of n returns: the 26 intraday returns plus the overnight. The overnight return bridges only consecutive trading sessions: when a stock's intraday record skips one or more sessions, the overnight return spanning that gap is omitted rather than computed across the hole. The coarse 15-minute frequency is deliberate: sampling coarsely mitigates the bid–ask bounce (Roll, 1984), nonsynchronous-trading, and Epps (1979) biases that contaminate correlation estimates at higher frequencies.

For the n daily returns, with the day index suppressed inside the sums, the annualized (by the trading-day count of a year) realized variance and covariance are

$$RV_i^d = 252 \sum_k r_{i,k}^2, \qquad RCov_{ij}^d = 252 \sum_k r_{i,k}\, r_{j,k}. \tag{1}$$

Their downside counterparts restrict the sum to negative returns (Bollerslev et al., 2020, 2022):

$$RV_i^{d-} = 252 \sum_k r_{i,k}^2 \mathbf{1}\{r_{i,k} < 0\}, \qquad RCov_{ij}^{d-} = 252 \sum_k r_{i,k}\, r_{j,k} \mathbf{1}\{r_{i,k} < 0\} \mathbf{1}\{r_{j,k} < 0\}. \tag{2}$$

$RCov^{d-}$ is the concordant-negative (both-down) semicovariance; pairwise sums run over the intersection of the two stocks' available intervals.

A realized correlation standardizes a covariance by the corresponding realized variances. Aggregating the daily measures to horizon h by the mean over a day window,

$$RCov_{ij}^h = \overline{RCov_{ij}^d}, \quad RV_i^h = \overline{RV_i^d}, \qquad RC_{ij}^h = \frac{RCov_{ij}^h}{\sqrt{RV_i^h\, RV_j^h}}, \tag{3}$$

and, for the negative semicorrelation, standardizing by the negative semivariances,

$$RC_{ij}^{h-} = \frac{RCov_{ij}^{h-}}{\sqrt{RV_i^{h-}\, RV_j^{h-}}}. \tag{4}$$

The numerator is the concordant-negative semicovariance defined above, aggregated to horizon h. Each horizon measure averages its daily counterpart over that measure's own available days: the (semi)covariance over the days both stocks trade, each (semi)variance over that stock's valid days. Being a correlation, RC^h is scale-free, so the annualization factor cancels from the ratio; but because the three averages need not share a common day count, RC^h can fall marginally outside $[-1, 1]$, and an "insanity filter" bounds it there, replacing any value above 1 (below -1) with 1 (-1) (Bollerslev et al., 2026, fn. 8); the same filter bounds every correlation feature, realized or implied, and the response.

The panel is monthly, one observation per pair per month, so the memory horizons are anchored at each month-end rather than being trailing daily windows: $h = d$ is the last trading day, $h = w$ the last five trading days, and $h = m$ all trading days of the calendar month (roughly 19–23 days; the fixed 22-day convention of daily-frequency HAR is not used). The response is the month-$(t{+}1)$ realized correlation, $RC_{ij,t+1} \equiv RC^m_{ij,t+1}$.

2.2. Factor-driven realized correlation features

A realized correlation estimated from a single month of data is noisy, and much of that noise is idiosyncratic. A factor structure offers a disciplined way to strip it out: if a factor model captures the common variation in returns well, the factor-driven component of return comovement is the systematic signal underlying the correlation, and the residual comovement is what a forecaster would rather discard. Because the factors themselves are not observed, we follow Fama and French (2020) in using observable firm characteristics as factor loadings to back the factor-driven correlations out of the realized covariances.

Let $\mathcal{N}_t$ denote the index members in month t and $N_t = |\mathcal{N}_t|$. Assuming the $N_t \times 1$ vector of returns is driven by K common factors, the realized covariance matrix at horizon h decomposes as $RCov^h = L\, RCov^h_f\, L' + RCov^h_\varepsilon$, where L is the $N_t \times K$ matrix of factor loadings, $RCov^h_f$ the $K \times K$ factor covariance matrix, and $RCov^h_\varepsilon$ the residual covariance matrix, whose off-diagonal elements carry the comovement the factors leave unexplained. Discarding that residual comovement while retaining the residual variances (so the diagonal continues to match that of $RCov^h$), and replacing the unobserved $RCov^h_f$ by $(L'L)^{-1}L'\, RCov^h\, L(L'L)^{-1}$ gives the factor-driven covariance matrix (Bollerslev et al., 2026, Eq. A.2)

$$RCov^{F,h} = P\, RCov^h\, P + \mathrm{Diag}\big(RCov^h_\varepsilon\big), \qquad P = L(L'L)^{-1}L', \tag{5}$$

with P the projection onto the span of the characteristics. Standardizing as in Section 2.1, and using that the diagonal of $RCov^{F,h}$ coincides with that of $RCov^h$ by construction, gives the factor-driven realized correlation

$$FRC^h_{ij} = \frac{\big[P\, RCov^h\, P\big]_{ij}}{\sqrt{RV^h_i\, RV^h_j}}, \qquad h \in \{d, w, m\}. \tag{6}$$

Equation (6) inverts only the $K \times K$ Gram matrix $L'L$, never $RCov^h$ itself, so the feature is well defined even though a month of data cannot identify a full $N_t \times N_t$ covariance matrix. The projection is not guaranteed to return a correlation inside $[-1, 1]$, and the insanity filter of Section 2.1 bounds it there. The three horizons give three features, FRC^d, FRC^w, and FRC^m. The matrix $RCov^h$ is assembled cross-sectionally each month from the same measures of Section 2.1 through $RCov^h_{ij} = RC^h_{ij}\sqrt{RV^h_i RV^h_j}$, so the factor-driven features rest on the same realized quantities the other models use.

The loadings are low-frequency firm characteristics, measured as of the end of month t so that $FRC^h_{ij,t}$ carries no look-ahead. Following Bollerslev et al. (2026), L collects $K = 14$ columns: the CAPM beta, size, book-to-market, and short-term reversal, together with ten of the eleven mispricing anomalies of Stambaugh, Yu and Yuan (2012) (financial distress, net stock issues, composite equity issues, accruals, net operating assets, momentum, gross profitability, asset growth, return on assets, and investment-to-assets), whose sources and construction are given in Section 3.[1] Because the characteristics differ widely in scale, each is rank-transformed cross-sectionally into $[-1, 1]$ each month (Gu, Kelly and Xiu, 2020), the CAPM beta excepted, which enters in its natural units. That exemption is not cosmetic. The rank transform centers a column, so a loading matrix assembled only from ranked characteristics spans a subspace essentially orthogonal to the constant vector; but the systematic covariance $\beta_i\beta_j\sigma^2_M$ points along β, whose cross-sectional mean is near one. Centering the beta as well would leave P annihilating the market component, the very source of the positive comovement that the correlations are made of, and FRC^h would collapse to a mean-zero series bearing no relation to RC^h. Left in levels, $P\beta = \beta$ and the systematic component passes through the projection untouched. Ranks are formed within $\mathcal{N}_t$, so a loading measures a stock's standing in the same universe over which the correlations are computed. A missing characteristic is set to the cross-sectional median, zero after the rank transform, rather than causing the stock to be dropped; a stock enters $\mathcal{N}_t$ only if its size and beta are observed and it has a realized variance at all three horizons, and any characteristic missing for the entire cross-section in a given month is dropped from L that month, keeping $L'L$ invertible.[2]

2.3. Option-implied correlation features

Option prices reveal the market's risk-neutral expectations; differencing off the physical, realized counterpart isolates the *risk premium* they embed, the forward-looking compensation for risk that backward-looking realized history cannot carry. This premium is the organizing idea of our implied features: a signal whose realized counterpart is cleanly measured enters as a premium (implied minus realized), and one without such a counterpart enters as a forward-looking level. Every implied feature is measured at the 30-day horizon, matching the one-month-ahead response, and

[1] The eleventh, the O-score, is distributed only in binarized form, as an indicator for the distressed tail rather than the score itself, and is constant across all but a fraction of a percent of our stock-months. As a factor loading it would add almost no variation to the span of L while risking an ill-conditioned $L'L$, so we omit it; financial distress, which enters as a continuous failure probability, already spans that dimension.

[2] The variance requirement binds only at $h = d$, whose window is the single last trading day, and costs on the order of one stock a month. It cannot be relaxed: P is dense, so a stock carrying no variance at a horizon would propagate through $P\, RCov^h P$ to every entry and void the whole cross-section rather than its own row. One cross-section serves all three horizons because the ranks are formed within a single $\mathcal{N}_t$.

taken as of the last trading day of month t, so, like the realized features, it carries no look-ahead. Matching the option maturity to the forecast horizon is deliberate: a longer-dated implied measure prices risk-neutral expectations over a window that overshoots the one-month response.

The construction rests on model-free per-stock risk-neutral inputs. The simple variance-swap rate SV_i (Martin, 2017) is the risk-neutral expected variance of the simple return and enters the correlation measures below as a ratio; the model-free implied variance $MFIV_i$ and skewness $MFIS_i$ of the log return (Britten-Jones and Neuberger, 2000; Bakshi, Kapadia and Madan, 2003) are its log-return counterparts, with $MFIV_i$ reserved for the same-basis variance premium; and the option-implied market beta $\beta_i^{\mathbb{Q}}$ (Buss and Vilkov, 2012), recovered from a risk-neutral variance–covariance matrix that pairs variance-swap rates with a parametrically corrected historical correlation matrix, is the forward-looking counterpart of a stock's market beta. Systematic risk is priced in single-name option prices themselves (Duan and Wei, 2009), which is the microfoundation for reading them as comovement signals; Chang, Christoffersen, Jacobs and Vainberg (2012) provide the principal alternative beta recovery, from risk-neutral skewness.

We begin with the core feature. Under a single-index representation of returns the correlation of two stocks is $\rho_{ij} = \beta_i \beta_j \, \sigma_M^2 / (\sigma_i \sigma_j)$; evaluated with risk-neutral inputs it gives a forward-looking, pair-specific implied correlation,

$$IC_{ij,t} = \frac{\beta_{i,t}^{\mathbb{Q}} \, \beta_{j,t}^{\mathbb{Q}} \, \sigma_{M,t}^2}{\sqrt{SV_{i,t} \, SV_{j,t}}}, \tag{7}$$

where $\sigma_{M,t}^2$ is the market's simple variance-swap rate. Equation (7) is the option-implied analogue of the realized correlation RC_{ij}^m and, alone among the features, varies across pairs *and* over time *and* looks forward. Its content beyond realized history is the *pairwise correlation risk premium*,

$$CRP_{ij,t} = IC_{ij,t} - RC_{ij,t}^m, \tag{8}$$

the part of the forward-looking implied correlation not already spanned by the month-t realized correlation that HAR carries: the pair's price of correlation risk. Its common counterpart uses the Driessen et al. (2009) option-implied *equicorrelation* $\overline{IC}_t$ (construction detailed in Driessen, Maenhout and Vilkov, 2021), the single correlation reconciling the index's risk-neutral variance with its constituents', together with the matched realized equicorrelation $\overline{RC}_t^{\mathrm{DMV}}$ from the same source, whose difference is the *market correlation risk premium*,

$$\overline{CRP}_t = \overline{IC}_t - \overline{RC}_t^{\mathrm{DMV}}. \tag{9}$$

The two premia are complementary: $\overline{CRP}_t$ carries the market-wide common component, $CRP_{ij,t}$ the cross-sectional tilt. The implied correlations and equicorrelations are bounded to $[-1, 1]$ by the insanity filter of Section 2.1.

One algebraic fact disciplines what the pairwise premium can and cannot claim, and we state it here rather than leave it to a referee. Every specification that carries CRP_{ij} also carries RC_{ij}^m among its realized features, and Equation (8) is a linear combination of the two; a least-squares regression containing both therefore fits exactly the model it would fit with IC_{ij} entered as a level, and produces identical forecasts. For the fixed specifications the premium form is a normalization rather than a restriction: it fixes what the coefficient measures (the loading on the forward-looking wedge, with the realized leg's offset absorbed by RC^m's own coefficient), but it adds no content, and no fixed-grid result below depends on differencing rather than entering the level. The form genuinely binds in two places. The market premium $\overline{CRP}_t$ differences off a realized series that appears nowhere else in any model, so there premium-versus-level is a substantive forward-looking claim; and the LASSO's ℓ_1 penalty is not invariant to reparametrization, so under selection the premium form acts as a prior rather than a relabeling.

The implied leg of this premium, however, is not history-free. The betas of Buss and Vilkov (2012) in Equation (7) are recovered from a risk-neutral variance–covariance matrix that pairs option-implied variances with a parametrically corrected *historical* correlation matrix, estimated from roughly a year of past returns. What enters Equation (8) as implied-minus-realized is therefore not option prices minus one month of realized history: the implied leg itself embeds long realized-correlation memory that the three-horizon HAR family does not carry, and a gain over a short-memory benchmark could in principle be that memory rather than anything the option market knows. Hollstein and Prokopczuk (2016) document the same structure from the measurement side: in their horse race across beta estimators the Buss–Vilkov construction is the best-performing of the option-implied betas precisely because its historical input acts as a

cross-sectional debiasing of a noisy risk-neutral quantity. We treat this as the paper's central identification problem and answer it from both directions. *By construction*, we rebuild the premium with the Martin and Wagner (2019) risk-neutral betas,

$$\beta_{i,t}^{\mathrm{MW}} = 1 + \frac{SV_{i,t} - \overline{SV}_t}{2\,SV_{M,t}}, \tag{10}$$

functions of date-t option prices alone: $\overline{SV}_t$ is the value-weighted average of the members' simple variance-swap rates, taken over the stocks whose rate and point-in-time capitalization are both observed that month (85% of the member cross-section in the average month), and $SV_{M,t}$ the market's rate. Substituting β^{MW} into Equation (7) gives $IC_{ij,t}^{\mathrm{MW}}$, and $CRP_{ij,t}^{\mathrm{MW}} = IC_{ij,t}^{\mathrm{MW}} - RC_{ij,t}^{m}$ is the correlation risk premium with the return history removed by construction; its market leg is the same $\overline{CRP}_t$, which is built from index and single-name variances and never touches a beta. Two properties make the substitution clean. The value-weighted mean of Equation (10) is exactly one, so the cross-section is centered without any historical anchor, the very service the historical input performs for Buss–Vilkov; and a member missing from the average shifts every beta by a common constant, which cannot re-rank pairs. *By control*, the fourth realized baseline of Section 2.4 hands the benchmark the same long memory directly, as trailing means of the pair's own realized correlation spanning the year the Buss–Vilkov leg is anchored on. The two arms bracket the question this section's features exist to answer: if the premium's forecasting content is the embedded history, it should vanish from CRP^{MW} and be absorbed by the long-memory benchmark; if it is genuinely forward-looking, it should survive both.

The remaining inputs are per-stock, and each enters the pairwise model through a symmetric summary whose form matches its channel: an additive risk-state *level* through the pair average, a comovement *loading* through the pair product, with the absolute difference added in each case to capture heterogeneity. The tables abbreviate these summaries as $\overline{X}$ for the pair average $\frac{1}{2}(X_i + X_j)$ and $|\Delta X|$ for the pair dispersion $|X_i - X_j|$; the market-wide series $\overline{IC}_t$ and $\overline{CRP}_t$ carry a time subscript only, which is what distinguishes them from the pair averages.

The per-stock variance risk premium $VRP_{i,t} = MFIV_{i,t} - RV_{i,t}^{m}$ differences two log-return variances, a clean same-basis premium and the measure whose aggregate counterpart Carr and Wu (2009) quantify and Bollerslev et al. (2009) establish as a return predictor; its realized leg RV_i^m is winsorized cross-sectionally each month at the 2.5% level, matching the winsorization the option-implied inputs already carry and taming a thin tail of realized-variance outliers. Being a risk-state level, it enters as the pair average $\frac{1}{2}(VRP_i + VRP_j)$ and dispersion $|VRP_i - VRP_j|$, gauging the priced variance risk, and hence the risk aversion, borne by the pair. Single-name variance premia are documented to be small relative to the index's (Bakshi and Kapadia, 2003a,b), a prior the grid's results will confirm.

Risk-neutral skewness has no cleanly measured realized counterpart at the monthly horizon, so it enters as a forward-looking level: the pair average $\frac{1}{2}(MFIS_i + MFIS_j)$ and dispersion $|MFIS_i - MFIS_j|$, a gauge of priced crash risk and return asymmetry. Single-name risk-neutral skewness is itself substantially systematic (Dennis and Mayhew, 2002), and tail risk carries much of the variance premium's compensation (Bollerslev and Todorov, 2011), so the pair summaries should be read as coarse gauges of the higher-moment channel rather than sharp ones.

Systematic covariance, finally, is bilinear in the market loadings ($\beta_i\beta_j\sigma_M^2$), so the implied betas enter through the product $\beta_i^{\mathbb{Q}}\beta_j^{\mathbb{Q}}$, the systematic-comovement channel, and the dispersion $|\beta_i^{\mathbb{Q}} - \beta_j^{\mathbb{Q}}|$, the loading heterogeneity that the single-index correlation does not span.

The pairwise correlation risk premium CRP_{ij} is the core forward-looking feature added to the realized benchmark, and CRP_{ij}^{MW} its history-free counterpart. The full implied block (both pairwise premia and the market premium, together with the variance-premium, implied-skew, and implied-beta features) is available to the penalized specification of Section 2.4, where selection guards against overfitting.

2.4. Forecasting models

Every model predicts the month-(t+1) realized correlation of a pair as a linear combination of features observed through month t,

$$\widehat{RC}_{ij,t+1} = b_0 + x_{ij,t}'b, \tag{11}$$

where $x_{ij,t}$ collects the predictors and the coefficients (b_0, b) are common to all pairs (Section 2.5). The competing models differ in which features enter $x_{ij,t}$; each nests the HAR benchmark, so a model's out-of-sample gain over HAR isolates the predictive content of the features it adds.

The benchmark is the heterogeneous autoregressive (HAR) regression of Corsi (2009), transplanted from realized volatility to realized correlation as in Audrino and Corsi (2010) and Bollerslev et al. (2026); it captures the persistent, long-memory-like dynamics of the realized measure through three horizons,

$$x_{ij,t}^{\text{HAR}} = \left(RC_{ij,t}^{d},\ RC_{ij,t}^{w},\ RC_{ij,t}^{m}\right)'. \tag{12}$$

The monthly term is the response's own operator lagged one month, $RC_{ij,t}^{m} \equiv RC_{ij,t}$: it uses only month-$t$ information, carries no look-ahead, and is the single strongest predictor of next month's correlation, making HAR a demanding benchmark rather than a naive one. We designate it *the* benchmark against which every competing model is judged out of sample; the comparison criteria are defined in Section 2.6.

The semivariance-HAR (SHAR) model augments HAR with the three downside semicorrelations of Section 2.1, in the nomenclature of Patton and Sheppard (2015),

$$x_{ij,t}^{\text{SHAR}} = \left(RC_{ij,t}^{d},\ RC_{ij,t}^{w},\ RC_{ij,t}^{m},\ RC_{ij,t}^{d-},\ RC_{ij,t}^{w-},\ RC_{ij,t}^{m-}\right)', \tag{13}$$

letting correlation dynamics differ in the wake of joint downside moves, when comovement is known to strengthen. HAR and SHAR are the two price-only baselines built entirely from the realized measures of Section 2.1.

The factor-driven model (SHAR-F) augments SHAR with the three factor-driven realized correlations of Section 2.2,

$$x_{ij,t}^{\text{SHAR-F}} = \left(RC_{ij,t}^{d},\ RC_{ij,t}^{w},\ RC_{ij,t}^{m},\ RC_{ij,t}^{d-},\ RC_{ij,t}^{w-},\ RC_{ij,t}^{m-},\ FRC_{ij,t}^{d},\ FRC_{ij,t}^{w},\ FRC_{ij,t}^{m}\right)', \tag{14}$$

offering the forecaster, alongside each pair's own noisy realized history, that history's projection onto the cross-section of firm characteristics.

The long-memory model (SHAR-F-L) closes the realized ladder. The monthly realized correlation is persistent well beyond the one-month window that is HAR's longest memory (Section 3), and Section 2.3 identifies roughly a year of that history sitting inside the implied leg of the correlation risk premium. SHAR-F-L therefore augments SHAR-F with trailing means of the pair's own monthly realized correlation,

$$x_{ij,t}^{\text{SHAR-F-L}} = \left(x_{ij,t}^{\text{SHAR-F}\prime},\ RC_{ij,t}^{3m},\ RC_{ij,t}^{6m},\ RC_{ij,t}^{12m}\right)', \qquad RC_{ij,t}^{\ell m} = \frac{1}{\ell}\sum_{s=0}^{\ell-1} RC_{ij,t-s}^{m}, \tag{15}$$

the windows spanning 3/6/12 months through the end of month t; a window with any month missing yields no feature rather than a partial average, so the panel's ragged edges are missing just where the history is genuinely incomplete. Extending the memory to a year makes this the by-control arm of the identification in Section 2.3: SHAR-F-L carries, as plain realized features, the same long correlation memory the premium's implied leg is anchored on, so whatever an implied feature adds *at this rung* cannot be that memory. HAR, SHAR, SHAR-F, and SHAR-F-L are our four realized baselines: each nests the one before it, and all four are backward-looking, using no option data.

The option-implied models augment a realized baseline with the forward-looking features of Section 2.3, one economic group at a time. Our main specification adds the correlation risk premium, its pairwise and common components together,

$$x_{ij,t}^{\text{HAR+CRP}} = \left(RC_{ij,t}^{d},\ RC_{ij,t}^{w},\ RC_{ij,t}^{m},\ CRP_{ij,t},\ \overline{CRP}_{t}\right)', \tag{16}$$

so that its R^2_{OOS} against HAR isolates the incremental content of the forward-looking correlation risk premium. A second premium group replaces the pairwise leg with its history-free counterpart, $\{CRP_{ij}^{\text{MW}},\ \overline{CRP}_{t}\}$: identical in every respect except the betas inside the implied correlation, which is what makes the pair of premium groups an instrument rather than a robustness check. Three further groups augment a baseline in turn with the variance risk premium $\{\frac{1}{2}(VRP_i+VRP_j),\ |VRP_i-VRP_j|\}$, the implied skewness $\{\frac{1}{2}(MFIS_i+MFIS_j),\ |MFIS_i-MFIS_j|\}$, and the implied betas $\{\beta_i^{\mathbb{Q}}\beta_j^{\mathbb{Q}},\ |\beta_i^{\mathbb{Q}}-\beta_j^{\mathbb{Q}}|\}$.

Each of the five implied groups augments each of the four realized baselines, giving a 4×5 cross of implied-augmented models alongside the four baselines themselves. The point of crossing them is to read the implied features' marginal value against a benchmark of increasing strength. An augmented model's R^2_{OOS} recomputed against its own

baseline (the baseline replacing HAR in the denominator of the evaluation criterion of Section 2.6) measures what an implied group contributes *at that baseline*; tracking that statistic across HAR, SHAR, SHAR-F, and SHAR-F-L shows how much of the forward-looking content survives as the realized measures are made progressively more demanding. A signal that adds to HAR but not to SHAR-F-L is one the realized history already spans.

Two penalized specifications close the grid, selecting features by the least absolute shrinkage and selection operator (LASSO): one over the twelve realized features (the three HAR correlations, their three downside semicorrelations, the three factor-driven correlations, and the three trailing means), and one over those twelve together with the full nine-feature implied block. Sharing the estimator and the realized core, the two differ only in the option-implied features, so the gap between them is a direct, selection-based estimate of the implied block's contribution. With the four baselines, the twenty augmented models, and this pair, the grid holds 26 models. Every model shares the linear form of Equation (11) and is estimated and evaluated identically.

2.5. Estimation

We estimate a single coefficient vector for the whole cross-section rather than one regression per pair. Following Bollerslev et al. (2026), the correlation dynamics are close to identical across pairs, so pooling sacrifices little, and the resulting panel of many thousands of pairs each month, millions of pair-month observations per estimation window, supplies the many observations that stabilize the coefficients and, for the larger feature sets, make feature selection reliable.

Two estimators share this pooled design. Ordinary least squares (OLS) fits a fixed predictor set. For the larger feature sets we also use the least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996), which adds an ℓ_1 penalty to the least-squares loss,

$$\min_{b_0,\,b} \frac{1}{|\mathcal{T}|} \sum_{(i,j,t)\in\mathcal{T}} \left(RC_{ij,t+1} - b_0 - x'_{ij,t} b\right)^2 + \lambda \sum_p |b_p|, \tag{17}$$

where $\mathcal{T}$ is the training sample and $\lambda \geq 0$ is the shrinkage parameter. At $\lambda = 0$ the estimator collapses to OLS; for $\lambda > 0$ it shrinks the slopes and sets some exactly to zero, selecting features from the full set. The intercept b_0 is left unpenalized. So that a single λ penalizes each slope comparably, we standardize every predictor by its mean and standard deviation before fitting; these moments are computed on the training window only, which also rules out look-ahead. The response is left in levels.

Forecasting is out of sample throughout, on a rolling five-year window refit once a year. For test year $t{+}1$, OLS estimates its coefficients on the pooled panel of the five preceding years $t{-}4, \ldots, t$ and forms one-month-ahead forecasts $\widehat{RC}_{ij,t+1}$ for every month of year $t{+}1$; the window then rolls forward one year and the models are re-estimated. LASSO uses the same five-year block but reserves its most recent year as a validation sample: it fits a path of models on years $t{-}4, \ldots, t{-}1$, chooses λ by prediction accuracy on the held-out year t, and then forecasts year $t{+}1$. Refitting annually lets the coefficients, and under LASSO the selected features, adapt as market conditions change. The initial window is consumed by estimation, so out-of-sample forecasts begin only once it has elapsed. The resulting one-month-ahead forecasts $\widehat{RC}_{ij,t+1}$ are the object of the out-of-sample evaluation developed next. Every forecast in this paper, and every test performed on one, is generated this way. Section 4.1 additionally quotes coefficients fit once on the whole sample; those estimates are descriptive, saying what the models load on, and no forecast, statistic, or claim in the paper is built from them.

2.6. Forecast evaluation

The one-month-ahead forecasts $\widehat{RC}_{ij,t+1}$ are evaluated out of sample on the pooled set $\mathcal{O}$ of pair-months for which a forecast is available: every pair-month beyond the initial estimation window of Section 2.5 whose predictors are observed for each competing model. Requiring the full predictor set present makes $\mathcal{O}$ common to all models, so differences in performance reflect predictive content rather than sample composition. The loss is the squared forecast error of the realized correlation, $(RC_{ij,t+1} - \widehat{RC}_{ij,t+1})^2$.

Accuracy is summarized by the out-of-sample R^2 of a model relative to the HAR benchmark of Equation (12) (Bollerslev et al., 2026),

$$R^2_{\text{OOS}} = 1 - \frac{\sum_{(i,j,t)\in\mathcal{O}} \left(RC_{ij,t+1} - \widehat{RC}_{ij,t+1}\right)^2}{\sum_{(i,j,t)\in\mathcal{O}} \left(RC_{ij,t+1} - \widehat{RC}^{\text{HAR}}_{ij,t+1}\right)^2}. \tag{18}$$

The benchmark is HAR rather than an unconditional or historical-mean forecast: HAR already carries the persistent correlation dynamics any useful model must reproduce, so R^2_{OOS} isolates the *incremental* content of the added features. HAR scores zero against itself, and a positive value means a model's pooled forecast-error sum of squares falls below HAR's. Because SHAR nests HAR, $R^2_{\text{OOS}}(\text{SHAR}) > 0$ precisely when the downside semicorrelations of Section 2.1 sharpen the out-of-sample forecast, the question this paper puts to each feature block.

Equation (18) weights every pair-month equally, and so measures accuracy over the cross-section of pairs rather than over invested capital. The distinction matters here because nothing guarantees a gain is size-neutral: a forecast that improves only on small, illiquid names is a different result from one that improves on the pairs a portfolio would actually hold. Bollerslev et al. (2026) put the question with a value-weighted R^2_{OOS}, weighting each pair-month by the product of its two firms' capitalizations. We ask it instead with the profile that weighted average cannot show. A product of capitalizations is a severely concentrated weight, placing most of its mass on pairs of the very largest firms, so it answers "how accurate are we on the biggest pairs?" while appearing to answer "how accurate are we, weighted by size?" Either way, a single number cannot say *where* across the size distribution a gain sits. We therefore report the equal-weighted R^2_{OOS} within terciles of pair size, defined by the average capitalization $\frac{1}{2}(ME_{i,t} + ME_{j,t})$ of the two firms and cut *within each month*, so that the terciles are cross-sectional and carry no time tilt: capitalizations grow several-fold over the evaluation period, and a pooled cut would sort the early sample into the small tercile and the late sample into the large one. Every stock in $\mathcal{O}$ has an observed capitalization by construction, since size is one of the two anchors a stock must satisfy to enter the month-t cross-section (Section 2.2), so the profile is computed on the identical sample as Equation (18) and differs from it in nothing but the partition (Section 6.2).

The statistical significance of a pairwise accuracy gain is assessed with the Diebold and Mariano (1995) test, DM hereafter, on the loss differential

$$d_{ij,t+1} = \left(RC_{ij,t+1} - \widehat{RC}^{\text{HAR}}_{ij,t+1}\right)^2 - \left(RC_{ij,t+1} - \widehat{RC}^{\theta}_{ij,t+1}\right)^2 \tag{19}$$

between the benchmark and model θ, so that $\mathbb{E}[d_{ij,t+1}] > 0$ signals that θ is the more accurate. The pooled differentials are strongly cross-sectionally dependent, because pairs share common market-wide shocks within a month, so a test treating pair-months as independent would be severely oversized. We therefore average $d_{ij,t+1}$ across pairs within each month and base the statistic on the resulting monthly series $\bar{d}_{t+1}$, with a Newey and West (1987) long-run variance and the small-sample adjustment of Harvey, Leybourne and Newbold (1997). This monthly series is the natural unit of independent information in a pooled panel, and the criterion by which every richer model is judged against HAR; the lag length follows the standard $\lfloor 4(T/100)^{2/9} \rfloor$ rule, 4 lags at our 168 months. It is also where we depart from Bollerslev et al. (2026), who instead average each pair's loss differential into its two constituent stocks and treat the resulting stock-level means as independent observations; that design ignores the common monthly shocks the pooled panel shares, and the very large test statistics it produces are the expected consequence.

Both statistics are also read within *subsamples* of the evaluation window: regimes of the market's price of variance in Section 4.2, calendar blocks in Section 6.1. Partition the response months into groups $g = 1, \dots, G$ and write $\mathcal{O}_g$ for the pair-months whose response month falls in group g. The conditional R^2_{OOS} is then Equation (18) taken over $\mathcal{O}_g$ in place of $\mathcal{O}$, and the conditional DM statistic is Equation (19) on that group's months alone.

Neither can establish that the groups *differ*, and reading them as if they could is an error of arithmetic rather than of judgment. The statistic is a time-series statistic on $\bar{d}_{t+1}$, so it scales with the square root of the number of months: a group one third the length of the sample carries roughly 58% of the full-sample statistic, and three groups that individually fail to reject are the *expected* outcome of an effect that is identical in all three. The comparison itself is therefore tested at full sample length, by regressing the monthly loss differential on group indicators,

$$\bar{d}_{t+1} = \sum_{g=1}^{G} \mu_g \, \mathbf{1}\{t+1 \in g\} + u_{t+1}, \tag{20}$$

whose coefficients are the groups' mean differentials. With a Newey–West covariance we test $H_0 : \mu_1 = \cdots = \mu_G$ by a Wald statistic referred to a χ^2_{G-1} distribution, and the directional comparison $\mu_G - \mu_1$ by its t-statistic. Every month enters both, so neither pays the power cost that splitting the sample imposes.

Equation (20) is written in the units of the loss differential, in which a group of months that are simply harder to forecast carries larger differentials for arithmetic reasons alone. Because that is the very confound a variance-regime

cut invites, we estimate it a second time on the differential expressed as a share of the loss it is taken against,

$$s_{t+1} = \bar{d}_{t+1} \big/ \bar{L}^{\,\theta_0}_{t+1}, \qquad \bar{L}^{\,\theta_0}_{t+1} = \frac{1}{|\mathcal{O}_t|} \sum_{(i,j)\in\mathcal{O}_t} \big(RC_{ij,t+1} - \widehat{RC}^{\,\theta_0}_{ij,t+1} \big)^2, \tag{21}$$

where θ_0 is the benchmark the differential is measured against and $\mathcal{O}_t$ the month-t cross-section of $\mathcal{O}$. Equation (21) is in the units of R^2_{OOS}, the share of the benchmark's error a model removes; Equation (20) is in the units of the DM test. We report both, because a scale chosen after seeing the answer is not a test.

One further test prices the grid as a whole. With 26 models scored against the same benchmark, some cell will always look best, and that cell's individual test cannot answer the question a reader of a model grid should ask: whether the *family* contains anything that beats HAR once the search across the family is accounted for. We therefore run a reality check in the sense of White (2000) over every non-HAR model's monthly loss-differential series against HAR. The statistic is the maximum *studentized* differential, in the form of Hansen (2005), whose studentization repairs the known size distortion that poorly performing models induce in the unstudentized maximum; its distribution is approximated by a moving-block bootstrap of the monthly series, in blocks of 12 months over 1000 draws. Section 4.1 reports the result whichever way it cuts.

3. Data

Our universe is the historical membership of the S&P 500, intersected with the names for which intraday data are available, yielding 826 members. Because firms enter and leave the index and their intraday records begin and end at different dates, the resulting panel is unbalanced by construction; we impose no balanced-panel restriction, matching the delisted-inclusive sample of Bollerslev et al. (2026) and avoiding the survivorship bias a point-in-time constituent list would introduce.

Realized measures are built from FirstRate Data intraday trade prices, split- and dividend-adjusted, sampled at the 15-minute regular-trading-hours grid of Section 2. Where the methodology requires the market's trading calendar (the overnight return bridges only consecutive market sessions, Section 2.1), we take it from the SPDR S&P 500 ETF (SPY), the most liquid index tracker, drawn from the same vendor and processed identically to the single-name series but held outside the estimation panel.

The option-implied features are obtained pre-computed from the publicly available datasets of Rehman and Vilkov (2020), Buss and Vilkov (2020), and Driessen, Maenhout and Vilkov (2020), which build them from OptionMetrics implied-volatility surfaces together with CRSP returns and distribute them winsorized. From these we take the simple variance-swap rates and model-free implied skewness (Rehman and Vilkov, 2012), the option-implied market betas (Buss and Vilkov, 2012), and the S&P 500 implied equicorrelation (Driessen et al., 2009); the market's simple variance-swap rate is that of the SPDR S&P 500 ETF drawn from the same source, consistent with the single-name variances.

The firm characteristics that form the factor loadings of Section 2.2 come from two sources. The primary source is the Open Source Asset Pricing panel of Chen and Zimmermann (2022), which supplies published cross-sectional predictors built from CRSP and Compustat following Green, Hand and Zhang (2013); from it we take the CAPM beta, book-to-market, and the ten mispricing anomalies of Section 2.2. The second source is FirstRate Data quarterly financial statements, which supply size, taken point-in-time from each firm's most recent quarterly report before month-end. The remaining loading, short-term reversal, is distributed by neither source and is constructed from our own daily returns.

The market series consumed by the economic applications come from two standard sources: the value-weighted market excess return, the risk-free rate, and the pricing factors from the Kenneth French data library, and the eight macroeconomic predictors of Welch and Goyal (2008) from Goyal's updated dataset.

The estimation sample is the pooled panel of pair-months. We impose no separate liquidity screen: membership in the S&P 500 is itself a large-capitalization, high-liquidity filter, and data quality is enforced instead through completeness requirements on the intraday record. A pair-month (i, j, t) enters the estimation sample when both stocks are index members in months t and $t+1$, the realized features are available in the feature month t, and the response monthly realized correlation is available in month $t+1$; every model in the grid is estimated and evaluated on this one common sample (Section 2.6). A two-year pre-sample buffer (intraday history retained from 2003) ensures that the longest-memory lags have sufficient history at the start of the sample. The resulting study sample spans response months from January 2005 to December 2023.

Table 1 summarizes the features of Section 2 over the final estimation sample: the 16,307,557 pair-months for which every feature is available. The features begin in January 2005, with the study sample itself, and each month's features predict the following month's realized correlation, so the first forecast month is February 2005 and the sample runs over 227 response months to December 2023. This is the sample every model in the grid is estimated on, its earliest years forming the first rolling estimation block and the remainder scored out of sample (Section 2.5), so the moments below describe the data the forecasts are actually built from rather than a wider panel containing rows no model ever uses.

Panel A collects the realized and factor-driven correlations. The realized-correlation patterns largely echo Bollerslev et al. (2026). The sample means are all positive, so paired stocks comove positively on average. The negative semicorrelations RC^{h-} exceed their unsigned counterparts RC^{h} at every horizon, consistent with comovement strengthening in joint downturns. Dispersion narrows and persistence rises as the memory window lengthens from a day to a month: among the unsigned correlations RC^{h}, the daily measure is the most dispersed and the least persistent at a one-month lag, and the monthly measure the least dispersed and the most persistent. That the monthly realized correlation is the most persistent of the three is why the monthly term dominates the HAR benchmark of Equation (12) and makes it a demanding one.[3]

The factor-driven correlations differ from the realized correlations they are projected from in ways that matter for what follows. FRC^{h} is centered below RC^{h} at every horizon and is right-skewed where the realized correlations are close to symmetric: projecting the realized covariance onto the K characteristic loadings of Section 2.2 discards the pair-specific component of comovement and retains only the part the factor structure can express. Unlike the realized correlations, which are correlations by construction, the projection is not confined to $[-1, 1]$ on its own, and the insanity filter of Section 2.1 leaves a visible mass at the boundary: 1.45% of the daily factor-driven values sit exactly at $+1$, against 0.53% and 0.28% at the weekly and monthly horizons, which puts FRC^{d}'s 99th percentile at the bound in Table 1. The clipping is almost entirely an upper-tail phenomenon: the lower bound accounts for only 16,446, 551 and 258 pair-months at the three horizons, which is why every FRC^{h} row reports a minimum of -1 while none of the three is anywhere near that bound in the body of its distribution. The realized correlations, by contrast, essentially never reach either bound. Most important, FRC^{h} is markedly more persistent than RC^{h} at every horizon and at every lag we report: the factor-driven correlations retain substantial autocorrelation at a year, where the realized correlations retain almost none. A regressor that is more persistent than the variable it targets is more informative about that variable's future than its own lag is, which is why SHAR-F, rather than HAR, is the benchmark the option-implied features must beat in Section 4.

Panel B's option-implied features confirm the premia they are built to isolate. The pairwise correlation risk premium CRP_{ij} is positive on average and persistent (implied correlation sits above realized, and the gap carries across months), and the market premium $\overline{CRP}$ shares its sign; the variance risk premium is likewise positive on average. Risk-neutral skewness is negative on average and for the large majority of pair-months, the familiar option-implied crash asymmetry, but not for all of them: the 75th percentile of $\overline{MFIS}$ is still negative while its extreme upper tail turns positive, so a minority of pair-months carry positive implied skewness. At a one-month lag the implied-beta product is the most persistent of the implied features.

The realized correlation and the realized variance display distinct dynamics. Figure 1 plots the twelve-month moving average of the cross-sectional means of the monthly realized correlation and realized variance: both rise sharply in the 2008 financial crisis and the 2020 COVID-19 episode, but the realized variance swings far more widely than the realized correlation, which remains comparatively stable. Figure 2 shows that the unconditional distribution of the monthly realized correlations is unimodal and widely dispersed across pairs and over time.

4. Results

We begin with the paper's question put in the form the design was built to answer. Every model of Section 2.4 is estimated as a single pooled coefficient vector on a rolling five-year window refit annually (Section 2.5) and scored, like every model in the paper, on the one common sample $\mathcal{O}$ of Section 3 (14,144,670 pair-months over 168 response months, 2010–2023), so that rows differ only in the features they carry. Table 2 reports the full 26-model grid, one panel per realized rung: each of Panels A–D opens with its rung, HAR through the long-memory SHAR-F-L, and stacks the

[3] We report autocorrelations at lags of 1, 3, 6, and 12 months, pooled within pair after removing each pair's own mean. Bollerslev et al. (2026) tabulate AR(1, 5, 21, 63); because their panel is also monthly, these are monthly lags, their values coinciding numerically with the trading-day counts of the day/week/month/quarter horizons. Our AR(1) estimates fall in the same range as theirs.

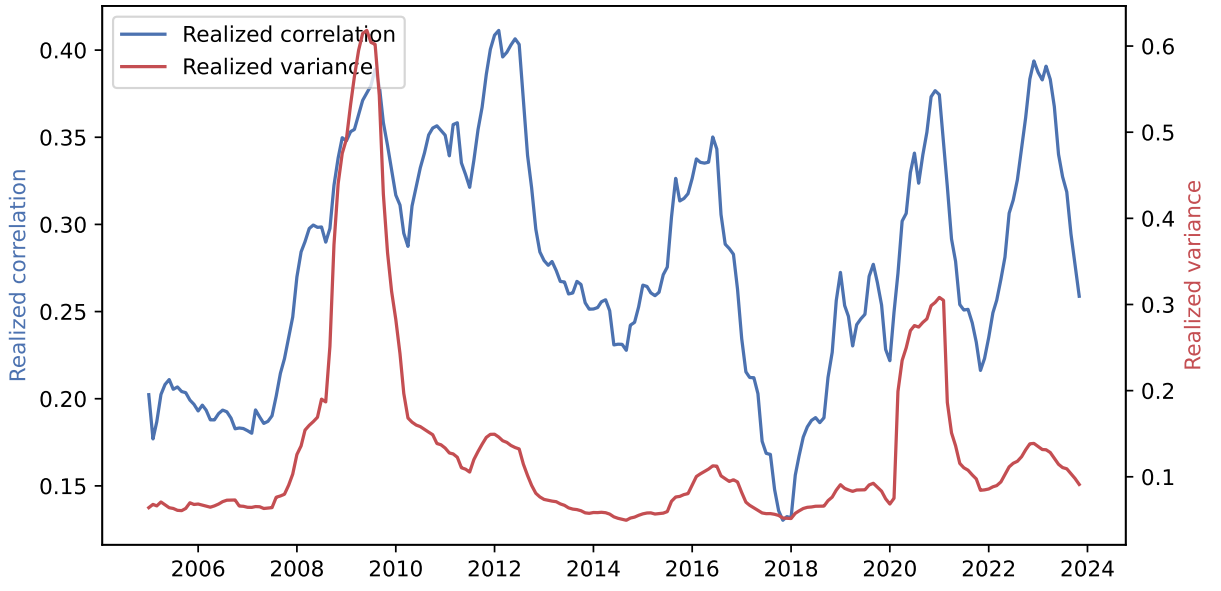

Figure 1: Monthly realized correlations and variances. Twelve-month moving average of the cross-sectional means of the monthly realized correlation (left axis) and realized variance (right axis) across the S&P 500 pairs in the final estimation sample, over the months January 2005 to November 2023.

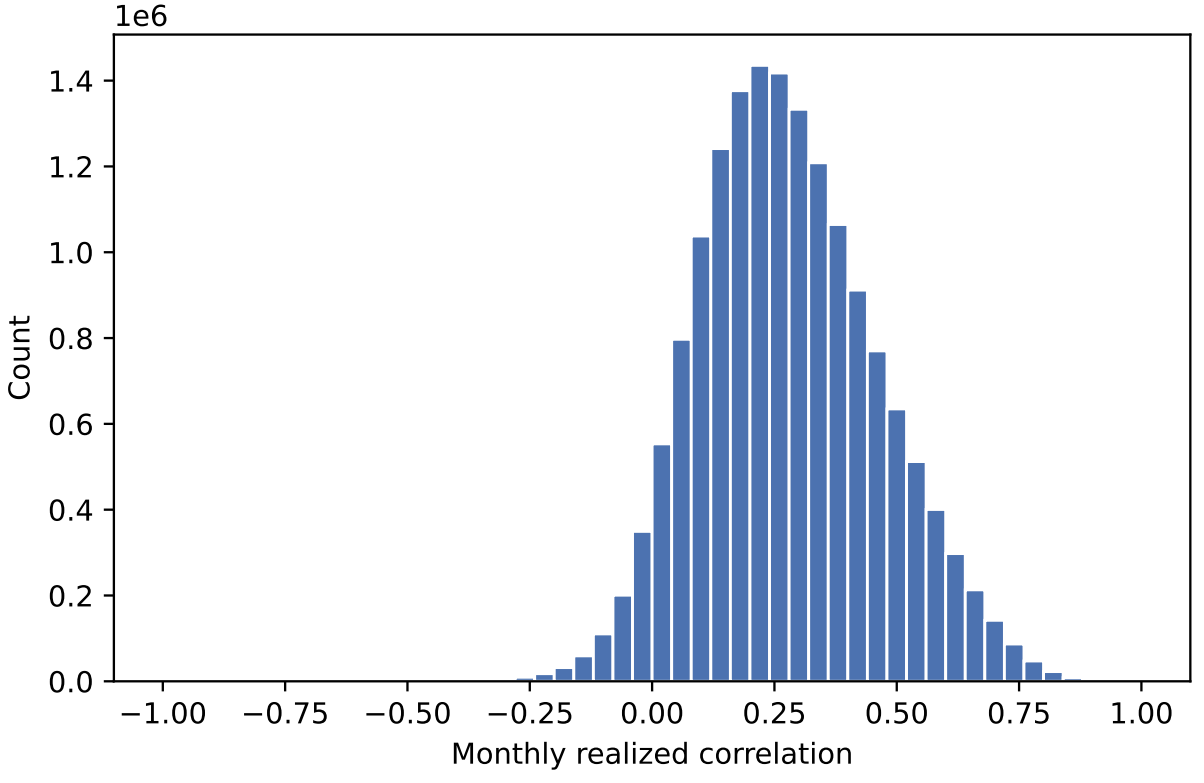

Figure 2: Distribution of monthly realized correlations. Histogram of the monthly realized correlation RC^m pooled over the 16,307,557 pair-months of the final estimation sample.

Table 1
Descriptive statistics for the realized-correlation, factor-driven, and option-implied features of the final estimation sample: the 16,307,557 pair-months over the 227 response months from February 2005 to December 2023 for which every feature is available. AR(k) is the within-pair autocorrelation at a monthly lag of k, pooled across pairs after removing each pair's own mean.

Variable	mean	std	skew	kurt	min	p1	p25	median	p75	p99	max	AR(1)	AR(3)	AR(6)	AR(12)
Panel A: Realized and factor-driven correlation features															
RC^d	0.29	0.30	-0.38	-0.29	-0.99	-0.46	0.09	0.32	0.52	0.86	1.00	0.06	0.08	0.03	0.04
RC^w	0.28	0.22	-0.01	-0.15	-0.98	-0.22	0.13	0.28	0.43	0.78	1.00	0.20	0.17	0.08	0.08
RC^m	0.28	0.18	0.24	-0.09	-0.88	-0.11	0.15	0.27	0.40	0.72	1.00	0.43	0.29	0.16	0.05
RC^{d-}	0.49	0.22	-0.02	-0.87	0.00	0.05	0.31	0.49	0.66	0.92	1.00	0.06	0.04	0.02	0.02
RC^{w-}	0.46	0.16	0.11	-0.42	0.00	0.12	0.34	0.46	0.58	0.84	1.00	0.13	0.13	0.06	0.10
RC^{m-}	0.45	0.14	0.25	-0.16	0.00	0.15	0.35	0.44	0.54	0.79	1.00	0.30	0.24	0.13	0.06
FRC^d	0.23	0.25	0.70	1.70	-1.00	-0.24	0.07	0.18	0.36	1.00	1.00	0.24	0.24	0.16	0.15
FRC^w	0.22	0.21	0.94	1.46	-1.00	-0.15	0.07	0.18	0.33	0.89	1.00	0.44	0.36	0.24	0.20
FRC^m	0.22	0.19	0.99	1.55	-1.00	-0.11	0.08	0.18	0.32	0.81	1.00	0.60	0.45	0.32	0.19
Panel B: Option-implied features															
CRP_{ij}	0.11	0.18	-0.24	0.58	-0.95	-0.35	-0.00	0.12	0.23	0.52	1.41	0.37	0.44	0.32	0.21
$\overline{CRP}$	0.03	0.12	-0.22	-0.17	-0.30	-0.27	-0.04	0.03	0.11	0.28	0.36	0.04	0.29	0.17	0.17
CRP_{ij}^{MW}	0.20	0.31	0.32	-0.20	-1.77	-0.43	-0.02	0.18	0.41	0.95	1.73	0.45	0.44	0.33	0.21
$\overline{VRP}$	0.03	0.12	-4.13	69.98	-4.24	-0.36	0.00	0.03	0.06	0.32	2.19	0.16	0.13	0.06	0.04
$\lvert\Delta VRP\rvert$	0.08	0.15	7.80	103.00	0.00	0.00	0.02	0.04	0.09	0.68	6.21	0.33	0.22	0.14	0.03
$\overline{MFIS}$	-0.65	0.38	-0.04	1.97	-2.78	-1.64	-0.86	-0.63	-0.43	0.35	2.93	0.30	0.22	0.13	0.06
$\lvert\Delta MFIS\rvert$	0.50	0.49	2.00	5.75	0.00	0.01	0.16	0.35	0.67	2.24	5.19	0.19	0.16	0.11	0.06
$\beta_i^{\mathbb{Q}}\beta_j^{\mathbb{Q}}$	1.12	0.60	0.87	1.87	-1.98	0.00	0.72	1.04	1.44	2.92	6.57	0.66	0.58	0.43	0.19
$\lvert\Delta\beta^{\mathbb{Q}}\rvert$	0.42	0.34	1.28	2.14	0.00	0.01	0.16	0.34	0.59	1.47	3.14	0.57	0.49	0.34	0.12

five implied groups on top of it, and Panel E closes with the penalized pair. Every model is scored twice. Against HAR, which puts the whole grid on one scale; and against its *own* baseline, the rung it augments, which tests the marginal contribution the text reads. The second score is the one that matters below, because a group adding nothing to a strong rung can still post a large HAR-relative statistic carried entirely by the rung's own gain. Read in that order, the table stages a decomposition: the option-implied gain is first established, then attributed, by construction and by control, to the ingredient that generates it.

4.1. The forecasting gain and its attribution

The gain is real. Augmenting HAR with the correlation risk premium (its pairwise and market components together, Equation (16)) raises the out-of-sample R^2 by 4.22% ($p = 0.015$): option-implied information does improve the monthly realized-correlation forecast of the benchmark realized model, and nothing below retracts that row. In sample, too, the premium enters with the sign the pricing mechanism of Section 2.3 predicts: a one-standard-deviation move in CRP_{ij} raises next month's realized correlation by 0.0305 ($t = 6.7$), and 92.4% of that loading survives the enlargement of the realized block to SHAR-F.[4] The question the rest of the table puts to it is not whether the gain exists but what it is made of, because Section 2.3 identified two candidates: information the option market holds about next month's comovement, and the year of realized-correlation memory that rides inside the premium's implied leg and that HAR, by construction, does not carry.

The attribution by construction is the third row of Panel A. HAR+CRP-MW differs from the main specification in one ingredient only: the betas inside the pairwise implied correlation are Martin–Wagner rather than Buss–Vilkov, functions of date-t option prices alone, with the market premium, the realized leg, and everything else held fixed. Strip the historical anchor out of the betas, in other words, and ask what forecasting content remains. The answer is none: 0.49% over HAR ($p = 0.767$), and the same null at every other rung of the grid. Whatever the premium forecasts with does not survive the removal of the one place return history enters its construction.

[4] Pooled OLS on standardized predictors with Driscoll and Kraay (1998) standard errors, fit once on the full sample; descriptive only. The full coefficient table is available upon request.

Table 2
Out-of-sample forecast evaluation of the full model grid. Equal-weighted out-of-sample R^2 relative to the HAR benchmark (Equation (18)) and the modified Diebold–Mariano statistic with its p-value (Equation (19)), for all 26 models of Section 2.4 over response months 2010–2023. Every model is estimated and scored on the one common sample of Section 3, so the rows differ only in which features enter. Panels A–D collect the four realized rungs, each opening with the rung itself and stacking on it the five implied groups (the correlation risk premium of the main specification HAR+CRP, its history-free Martin–Wagner counterpart CRP-MW, the variance risk premium, the implied skewness, and the implied betas). Panel E holds the penalized LASSO models, over the twelve realized features alone and over those together with the full nine-feature implied block, with the penalty selected on a held-out validation year. The left block scores every model against HAR, which puts the grid on one comparable scale. The right block scores each model against its *own* baseline: the realized rung an implied group augments, and LASSO (realized) for LASSO (realized+implied), whose only difference is the implied block. It is empty for the four realized rungs and for the models built on HAR, whose own baseline *is* HAR and whose test therefore already appears in the left block.

	vs. HAR			vs. own baseline		
Model	R^2_{OOS} (%)	DM stat	DM p	ΔR^2 (%)	DM stat	DM p
Panel A: HAR						
HAR	0.00					
HAR+CRP	4.22	2.45	0.015			
HAR+CRP-MW	0.49	0.30	0.767			
HAR+VRP	-2.87	-1.04	0.302			
HAR+skew	-0.16	-1.14	0.257			
HAR+$\beta^{\mathbb{Q}}$	1.78	4.56	0.000			
Panel B: SHAR						
SHAR	0.54	1.46	0.147			
SHAR+CRP	4.25	2.45	0.015	3.73	2.34	0.020
SHAR+CRP-MW	0.57	0.35	0.725	0.03	0.02	0.987
SHAR+VRP	-1.85	-0.77	0.443	-2.41	-1.06	0.289
SHAR+skew	0.38	0.96	0.337	-0.16	-1.03	0.303
SHAR+$\beta^{\mathbb{Q}}$	2.21	3.65	0.000	1.67	4.64	0.000
Panel C: SHAR-F						
SHAR-F	1.13	2.29	0.023			
SHAR-F+CRP	4.41	2.36	0.019	3.32	2.05	0.042
SHAR-F+CRP-MW	1.29	0.78	0.439	0.17	0.10	0.922
SHAR-F+VRP	-1.54	-0.51	0.611	-2.70	-1.07	0.288
SHAR-F+skew	0.93	1.66	0.099	-0.20	-1.24	0.217
SHAR-F+$\beta^{\mathbb{Q}}$	2.42	3.69	0.000	1.31	3.88	0.000
Panel D: SHAR-F-L						
SHAR-F-L	7.94	3.13	0.002			
SHAR-F-L+CRP	7.24	2.90	0.004	-0.76	-0.38	0.701
SHAR-F-L+CRP-MW	5.92	2.36	0.019	-2.20	-1.99	0.048
SHAR-F-L+VRP	6.79	2.43	0.016	-1.25	-0.74	0.459
SHAR-F-L+skew	7.63	2.92	0.004	-0.33	-1.59	0.113
SHAR-F-L+$\beta^{\mathbb{Q}}$	8.45	3.31	0.001	0.56	1.87	0.064
Panel E: LASSO						
LASSO (realized)	9.25	3.96	0.000			
LASSO (realized+implied)	9.21	4.11	0.000	-0.04	0.22	0.829

The attribution by control sits in Panel D, and it closes the case from the opposite side. SHAR-F-L extends the realized benchmark's memory to the same twelve months the premium's implied leg is anchored on (trailing means of the pair's own monthly realized correlation, Equation (15), realized features and nothing else) and beats HAR by 7.94% on its own (DM = 3.13, $p = 0.002$), nearly double the main specification's gain. At that rung the premium has nothing left to add: SHAR-F-L+CRP moves the forecast by −0.76% ($p = 0.701$), and the history-free premium is

significantly harmful there (-2.2%, $p = 0.048$). The two arms of the identification agree, and on the sharper of the two available readings: the premium's forecasting content over HAR is the long realized-correlation memory its implied leg happens to carry, supplied more cheaply, and more accurately, by the pair's own trailing year.

The realized ladder that conclusion stands on runs down the opening rows of the four panels, and it deserves a look of its own because the long-memory rung is not an incremental one. SHAR delivers a small, insignificant gain over HAR (0.54%), echoing Bollerslev et al. (2026), for whom SHAR likewise improves little on HAR alone. The factor-driven SHAR-F roughly doubles that gain (1.13%, $p = 0.023$): projecting the realized covariance onto the cross-section of firm characteristics recovers systematic signal a pair's own short history leaves behind. The trailing means then add more than both together, taking the ladder to 7.94% ($p = 0.002$). The monthly realized correlation is a highly persistent process (Section 3), and three plain averages of its own past turn out to carry more month-ahead information than the semicorrelations and the characteristic projection combined. That is the fact the premium's headline gain was standing on.

A design that stopped one rung short would have concluded the opposite, which is why the ladder matters. Against SHAR-F, the strongest benchmark without the trailing means, 78.6% of the premium's HAR-relative contribution survives, and SHAR-F+CRP still improves on its own baseline at the five-percent level ($p = 0.042$). Read there, the characteristic projection and the option market's correlation view look complementary. That is the conclusion an earlier draft of this paper drew, and the one a referee armed with the long-memory benchmark would have overturned. What SHAR-F lacks is not forward-looking information but memory; hand the benchmark the trailing year and the complementarity dissolves. Retention against the strongest realized benchmark is zero.

The penalized pair of Panel E reaches the same answer under selection rather than specification. A LASSO selecting over the twelve realized features attains $R^2_{\text{OOS}} = 9.25\%$, the largest gain in the grid; allowed to select over the same realized core *and* the entire nine-feature implied block (both premia, both legs of every group), it moves by -0.04 percentage points ($p = 0.829$). Given every implied feature and the freedom to select among them, the estimator finds nothing on the implied side worth carrying that the realized side does not already span.

The remaining implied groups fill in the pattern. The variance-risk-premium and implied-skew groups add nothing at any rung: their marginal contributions are negative throughout, so whatever the premia on variance and higher moments price, they carry no monthly correlation forecast at any level of realized sophistication. The implied betas are the one implied group whose marginal contribution remains positive at the top rung (0.56% over SHAR-F-L, $p = 0.064$): suggestive and not established, and we read it as no more than that. One reading of that residual can, however, be ruled out by measurement: that it is more of the memory this section has been dissolving, carried this time in historical betas rather than historical correlations. The physical-beta placebo of Section 5.1, the same pair features rebuilt from sixty-month *physical* market betas, contributes significantly at the weaker rungs (+2.05% over SHAR, +1.09% over SHAR-F) and adds 0.01% at this one (DM $= -0.11$, $p = 0.916$): the trailing means absorb the physical product entirely, as they absorb the premium, while the implied version's margin survives them. Whatever the implied betas still carry at the top rung, it is not historical beta structure.

The family-level check of Section 2.6 prices the whole search at once: the largest studentized statistic in the family is 4.57, at the implied-beta cell on HAR (HAR+beta), with a bootstrap p-value of 0.182. Priced as a family, the grid contains nothing that beats HAR at conventional levels. That is the conservative form of the statement, reported deliberately; the paper's conclusions do not depend on it either way, resting as they now do on two nulls and an absorption, none of which a multiplicity correction threatens.

The calendar profile localizes the channel from a third direction, and we volunteer it here rather than leave it to the robustness section. Across the sample's three calendar blocks the main specification's contribution over HAR falls monotonically, from 8.33% in 2010–2014 to 1.68% in 2020–2023, while the implied-beta group runs the other way, from 0.48% to 1.88% (Section 6.1). Whatever is fading is the priced-correlation premium specifically, not option-implied information at large.

What survives these attributions is not a forecasting margin at all. The premium's distinctive, option-specific property in this paper is structural: it makes the $N(N-1)/2$ separate forecasts a covariance matrix is assembled from cohere with one another, where a placebo with the same forecasting content actively fails to (Section 5.1). The applications of Section 5 open with that result.

4.2. When does the premium help?

The economics behind the correlation risk premium make a prediction about *when* it should help. If CRP carries information because the option market prices correlation risk (Driessen et al., 2009; Buss et al., 2019), its forecasting

content should be largest when that risk is most expensive: when the market's own price of variance is high. The attribution of Section 4.1 predicts the opposite: a signal whose content is realized memory has no reason to concentrate where variance is dear. We test the pricing logic's prediction on the forecasting margin rather than in the applications of Section 5: a third of our window can carry a statement about forecast accuracy, where the same months of portfolio returns cannot carry one about alpha or about a fee.

The conditioning variable is the S&P 500 30-day model-free implied variance, the same σ_M^2 that enters the implied correlation of Equation (7), observed at the end of the *feature* month, so that a month's regime is known when its forecast is made. Response months are sorted into terciles of it, 56 months each, with breakpoints cut on the full evaluation window and disclosed as such: the regimes say which months resemble each other, they are not a rule a forecaster could have followed. In implied-volatility units the cut points are 15.8% and 19.4%, and the three regimes average 14.0%, 17.3% and 25.7%. They are level cuts rather than calendar splits, and the sample does not sort itself neatly into eras: the high-variance regime arrives in 13 separate runs of consecutive months, the low in 17 and the middle in 28.

The pricing prediction is not confirmed. In Table 3 the main specification's contribution over HAR is 4.94%, 3.01% and 4.68% across the three regimes, against 4.22% over the whole window: no concentration in turbulent months, and no ordering in the price of variance. The per-regime DM tests reject in none of them ($p = 0.108$, 0.19 and 0.142), which is what a uniform effect looks like at a third of the sample (the arithmetic of Section 2.6) and is not evidence that the gain has gone anywhere.

The comparison itself is tested on all 168 months by Equation (20), and the answer depends on the scale, so we give both. In the units of the R^2 columns (the share of HAR's own error the premium removes, Equation (21)), the high-minus-low difference is negative and significant at five percent ($t = -2.14$, $p = 0.034$; Wald $p = 0.055$): the *proportional* gain is larger in calm months than in turbulent ones. In the units of the loss differential it is indistinguishable from zero ($t = 0.34$; Wald $p = 0.895$). The two disagree because turbulent months are harder to forecast for every model in the grid: the premium removes a smaller share of a larger error, and in levels the two effects cancel. The pooled R^2 columns sit between them, since pooling squared errors within a regime weights that regime's most turbulent months most heavily.

We read this as a non-confirmation, and as the conditional fingerprint of the attribution. The pricing direction fails, and what appears in its place, a gain that is proportionally *larger* in calm months, is significant on one of the two scales, shared by the other correlation-premium rows as near-identical specifications must share it, and anticipated by no pricing argument we know; we report it and build nothing on it. It is, however, just how the long-memory reading behaves: SHAR-F-L's own gain is likewise proportionally largest in the calm regime (Table 3), as it should be for a signal made of persistence: persistence forecasts best when the process is not being shocked. What the cell also establishes is a robustness the gain needs: the 4.22% is not a crisis artifact earned in a handful of turbulent months a reader is entitled to suspect. It is spread across the sample's variance regimes, and if anything thinner where the market's price of variance is highest.

Two rows frame the main result from either side. The implied-beta block is the grid's most regime-invariant: it contributes 2.08% and 1.94% at the two extremes, is significant in every regime ($p = 0.004$ in the dearest), and its contrast is flat ($t = -0.42$); the forward-looking loadings predict comovement in the same measure whatever variance costs. And the one row whose gain does sit in the turbulent regime is realized rather than implied: SHAR-F improves on HAR by 2.51% there ($p = 0.009$) against 0.67% in the calm regime, though on the equal-weighted monthly scale that difference fails to establish itself ($t = -0.38$). The block that behaves like a crisis instrument, in other words, is the factor-driven realized one, not the option-implied premium.

Three caveats travel with the table. The breakpoints are cut on the full window, so the regime map uses information the forecaster did not have; it is a description of the sample, and we do not report a conditional strategy that would need it in advance. A regime's months are not contiguous, and the long-run variance behind each per-regime DM statistic treats them as though they were. That is why the run counts above are reported rather than assumed away, and why the reading rests on the two full-sample tests, which are computed on the unbroken monthly series. Finally, the Wald statistic is mildly oversized at this sample length in simulation, so it is the directional contrast, not the omnibus, that carries the conclusion. We do not cut the applications of Section 5 this way at all: a regime cell of 56 monthly returns cannot establish an alpha or a fee, and reporting one would invite the very reading the power arithmetic above forbids.

Table 3
Out-of-sample forecast evaluation within regimes of the market's price of variance. Each model's marginal contribution and its modified Diebold–Mariano p-value within terciles of the S&P 500 30-day implied variance, over the 168 response months of the common sample of Section 3. The tercile is cut on the *feature* month, so a month's regime is known when its forecast is made, and on the full window (56 months per regime; cut points 15.8% and 19.4% in implied-volatility units). The contribution is measured against each model's own baseline, as in the right-hand block of Table 2: HAR for the realized rungs and for the models built on HAR, the augmented rung otherwise, and LASSO (realized) for LASSO (realized+implied). Every row therefore reports the quantity Section 4.1 reads; HAR is the benchmark and has no row, and the rows group into the same family panels as Table 2. Within a regime the R^2 pools squared errors and so weights that regime's most turbulent months most heavily. The final two columns are the high-minus-low contrast of Equation (20), estimated on all 168 months at once and reported as a t-statistic on both scales: *share*, the differential as a fraction of the baseline's own loss (Equation (21)), which is the scale of the R^2 columns, and *loss*, the raw differential, which is the scale of the DM columns. The per-regime tests use 3 Newey–West lags and the full-sample contrasts 4; a regime's months are not contiguous, and Section 4.2 reports how many runs each arrives in.

	Low MFIV		Mid MFIV		High MFIV		High−Low t	
Model	ΔR^2 (%)	DM p	ΔR^2 (%)	DM p	ΔR^2 (%)	DM p	share	loss
Panel A: HAR								
HAR+CRP	4.94	0.108	3.01	0.190	4.68	0.142	-2.14	0.34
HAR+CRP-MW	2.90	0.243	0.83	0.737	-1.85	0.691	-1.81	-0.89
HAR+VRP	1.38	0.003	0.39	0.498	-9.37	0.162	-1.78	-1.66
HAR+skew	-0.14	0.286	-0.32	0.310	-0.04	0.984	0.33	0.64
HAR+$\beta^{\mathbb{Q}}$	2.08	0.005	1.31	0.032	1.94	0.004	-0.42	0.27
Panel B: SHAR								
SHAR	-0.10	0.841	0.70	0.283	0.95	0.213	0.10	1.13
SHAR+CRP	4.23	0.137	2.76	0.190	4.17	0.155	-2.04	0.37
SHAR+CRP-MW	2.21	0.335	0.71	0.785	-2.44	0.566	-1.69	-0.93
SHAR+VRP	1.23	0.004	0.24	0.566	-7.88	0.148	-1.83	-1.74
SHAR+skew	-0.14	0.395	-0.37	0.258	0.01	0.832	0.11	0.75
SHAR+$\beta^{\mathbb{Q}}$	1.91	0.006	1.26	0.036	1.84	0.001	-0.24	0.35
Panel C: SHAR-F								
SHAR-F	0.67	0.370	0.02	0.895	2.51	0.009	-0.38	1.59
SHAR-F+CRP	3.88	0.192	3.09	0.171	3.04	0.288	-2.08	0.17
SHAR-F+CRP-MW	2.66	0.276	1.03	0.688	-2.78	0.535	-1.80	-1.04
SHAR-F+VRP	1.25	0.007	0.12	0.701	-8.68	0.162	-1.74	-1.66
SHAR-F+skew	-0.15	0.444	-0.48	0.184	0.00	0.925	-0.05	0.65
SHAR-F+$\beta^{\mathbb{Q}}$	1.45	0.019	1.12	0.047	1.36	0.013	-0.12	0.27
Panel D: SHAR-F-L								
SHAR-F-L	13.79	0.001	1.72	0.632	8.55	0.021	-2.36	-0.37
SHAR-F-L+CRP	-1.16	0.278	-0.35	0.833	-0.84	0.833	-0.80	0.20
SHAR-F-L+CRP-MW	-0.38	0.439	-0.99	0.641	-4.80	0.051	-1.42	-1.72
SHAR-F-L+VRP	0.93	0.036	0.46	0.417	-4.62	0.204	-1.88	-1.63
SHAR-F-L+skew	0.02	0.912	-0.20	0.308	-0.75	0.179	-0.32	-1.27
SHAR-F-L+$\beta^{\mathbb{Q}}$	0.58	0.069	0.84	0.091	0.26	0.612	-0.48	-0.35
Panel E: LASSO								
LASSO (realized)	13.64	0.000	7.20	0.050	7.36	0.061	-2.44	-0.64
LASSO (realized+implied)	0.30	0.985	2.17	0.062	-2.28	0.300	-0.84	-0.77

5. Economic value

Section 4 attributed the forecasting gain to realized memory. This section asks what is left for option prices to do, on the four margins where a correlation forecast reaches a portfolio, all four taken from Bollerslev et al. (2026): inverted into minimum-variance weights, collapsed into an aggregate predictor, read as a cross-sectional ranking, and asked to price a fixed position. All four applications evaluate the very models of Section 2.4, on the common evaluation sample

$\mathcal{O}$ of Section 2.6, with HAR+CRP the main specification throughout and the rest of the grid read for pattern rather than as 26 separate hypotheses; no application admits a model-free implied *level*, since Section 2.3 locates the design's forward-looking content in a premium. Every portfolio return in this section is net of the ten-basis-point one-way charge on realized turnover, the paper's standing cost assumption, applied wherever a comparison could turn on it. We lead with the section's positive finding, because it is the one property in this paper that option prices supply and no realized substitute does: the premium repairs the *coherence* of the covariance matrix assembled from the pairwise forecasts. The four uses then return one verdict on value, no fee a mean-variance investor would pay, and each null carries its mechanism.

5.1. Minimum-variance portfolios: incoherence and repair

The global minimum-variance (GMV) portfolio is the standard testing ground for a covariance forecast (Fleming, Kirby and Ostdiek, 2003): its weights depend on the covariance matrix alone, so it isolates the correlation model from expected returns. A one-month-ahead covariance forecast combines each model's correlation matrix with a common set of variance forecasts,

$$\widehat{RCov}^{\theta}_{t+1} = \widehat{D}_{t+1}\,\widehat{RC}^{\theta}_{t+1}\,\widehat{D}_{t+1}, \qquad \widehat{D}_{t+1} = \operatorname{diag}\big(\widehat{RV}^{1/2}_{1,t+1}, \dots, \widehat{RV}^{1/2}_{N,t+1}\big), \tag{22}$$

where $\widehat{RV}_{i,t+1}$ is a pooled log-HAR forecast of each stock's monthly realized variance, shared by every specification so that differences across Equation (22) reflect the correlation model alone. The GMV weights are

$$w^{\theta}_{t+1} = \frac{(\widehat{RCov}^{\theta}_{t+1})^{-1}\mathbf{1}}{\mathbf{1}'(\widehat{RCov}^{\theta}_{t+1})^{-1}\mathbf{1}}, \tag{23}$$

formed each month on the common set of stocks for which every model supplies a complete forecast matrix, held over month $t+1$, and evaluated on realized portfolio variance, the criterion the GMV optimizes, through the paired monthly variance differential of each model against its *own* baseline, put through the modified Diebold–Mariano statistic of Section 2.6 in the spirit of Engle and Colacito (2006) (Newey–West, 3 lags). Economic value is the Fleming et al. (2003) fee Δ_γ a mean-variance investor with risk aversion $\gamma \in \{1, 10\}$ would pay to switch from the HAR covariance, computed on returns net of the turnover charge for both portfolios,

$$U(r_t, \gamma) = (1 + r_t) - \frac{\gamma}{2(1+\gamma)}(1 + r_t)^2, \qquad \sum_t U(r^{\mathrm{HAR}}_t, \gamma) = \sum_t U(r^{\theta}_t - \Delta_\gamma, \gamma). \tag{24}$$

A matrix assembled pair by pair from a pooled forecast need not be positive definite, and frequently is not: nothing in Equation (11) constrains $\binom{N}{2}$ separately fitted values to cohere as a single Gram matrix. The final column of Table 4 reports how often each model's *raw* assembled matrix has a negative eigenvalue. HAR's never does; SHAR's fails in 14.9% of months, SHAR-F's in 38.7%, and the variance-risk-premium models' in 44.6% to 66.1%. Before any portfolio is formed, every matrix is corrected by the convexity approach of Bollerslev et al. (2026): $\widehat{RC}^{\theta}$ is shrunk toward the identity (Ledoit and Wolf, 2004) by the smallest weight that holds the condition number below 10,000, identically for every specification. Non-PD therefore describes the forecast that entered the correction, not the matrix that left, and every repair claim below is conditional on that regularization. Incoherence is not a technicality; it prices. Across the 26 models the failure frequency and the realized risk of the resulting portfolio have a Spearman rank correlation of 0.88, and the realized ladder runs backwards here: SHAR-F, a significantly better pairwise forecaster than HAR, builds a portfolio carrying 42% more realized risk (13.56% against 9.55%; DM $= -4.72$) on twice the turnover (10.04 against 5.02). Enriching the feature set sharpens each forecast and makes the forecasts jointly less consistent; the correction then shrinks the incoherent matrix toward the identity, which discards the very correlation structure the portfolio needed. A better correlation forecast and a better covariance matrix are different objects.

Against that backdrop, one implied group does something no realized feature set in the table does: adding the correlation risk premium *repairs* the matrix. On SHAR the raw failure rate falls from 14.9% to 11.9%, on SHAR-F from 38.7% to 25%, with realized volatility falling in step (11.21% to 10.30% and 13.56% to 12.02%); those two falls are the unconstrained table's only significant variance reductions, DM $= 3.76$ on SHAR and 2.71 on SHAR-F. The reduction is largest where there is most incoherence to repair: largest at SHAR-F, smaller at SHAR, unresolved at HAR, whose raw matrix never fails. Its consistency, meanwhile, is nearly identical across the three (82.7%, 84.5% and 83.9%

of months below baseline). This is what a signal carrying genuine *common* structure should do (Engle and Kelly, 2012; Herskovic, Kelly, Lustig and Van Nieuwerburgh, 2016): the premium's market-wide component disciplines the $\binom{N}{2}$ forecasts toward mutual consistency, and coherence is what a covariance matrix is made of.

Is the repair a property of option prices, or would any rank-one structure do? The question decides whether the result is a finding about the option market or about matrix algebra, and the placebo of Section 2.4's grid answers it by measurement. We rebuild the implied-beta pair features from the sixty-month *physical* market betas the paper already carries, the characteristic that anchors the loading matrix of Section 2.2 and the market-neutral constraint below, and estimate SHAR+β-P, SHAR-F+β-P, and, for the forecasting reading of Section 4.1, SHAR-F-L+β-P beside the grid, on the same sample, never inside it. In *forecasting*, the physical product is as good as the implied one: it adds 2.05% over SHAR ($t = 4.87$) and 1.09% over SHAR-F ($t = 2.1$), consistent with Section 4, since the physical beta product carries the same realized comovement memory. In *coherence* it is actively destructive: the placebo raises SHAR's failure rate from 14.9% to 26.2% where the premium cuts it to 11.9%, a gap of 14.3 percentage points at the same rung, and its realized-variance differential is significantly *negative* (DM $= -2.66$, $p = 0.008$) where the premium's is +3.76; at SHAR-F the failure gap is 18.5 points. Same forecasting content, same rank-one form, opposite verdicts on the matrix.

The repair is therefore not a generic property of adding rank-one beta structure, and not a property of the realized memory the forecasting gain reduced to: it needs the option-implied construction. It needs its historical anchor too, since the history-free CRP-MW, date-t option prices alone, breaks the matrix wherever it enters (Table 4). What repairs is the premium as the option market's correlation view is actually assembled, risk-neutral variance structure laid over the pair's realized correlation pattern, with the option content necessary for the repair rather than sufficient. That distinction is the two-margin contrast this paper adds to the forecasting result: the same placebo that matches the premium where its content is realized memory fails it on the one margin where the option market contributes something of its own.

Table 5 sharpens the repair into a portfolio result. A minimum-variance portfolio's risk is dominated by the market factor, which no correlation forecast can move, so we re-run the identical grid with the weights additionally constrained to zero market beta,

$$w_{t+1}^{\theta,\mathrm{BN}} = \arg\min_{w} \; w' \widehat{RCov}_{t+1}^{\theta} w \qquad \text{subject to} \qquad w'\mathbf{1} = 1, \quad w'\beta_t = 0, \tag{25}$$

with β_t the sixty-month rolling market betas of Section 2.2: an *exogenous* beta, identical across the grid, where the constraint of Bollerslev et al. (2026) depends on each model's own forecast. The cost of the choice is that our portfolios are market-neutral only to the accuracy of the beta estimate, and our fees are not directly comparable with theirs. The constraint is nearly free for the benchmark (HAR's volatility rises to 101% of its unconstrained level) and takes 1.71 percentage points off SHAR-F, whose errors run disproportionately in the market direction. With the common factor out of both sides of every comparison, the premium's edge is measurable everywhere: all three cells that add CRP to a realized rung reduce variance significantly (DM $= 2.33$ on HAR, 4.3 on SHAR, 3.78 on SHAR-F), where unconstrained the HAR cell was not (DM $= 0.99$, $p = 0.325$, on a 0.23-point margin 168 months cannot resolve). The honest statement of the mechanism is therefore that the premium's common structure reduces realized portfolio risk generally, by most where the realized models' own richness has broken the matrix, and that the unconstrained HAR cell was a question of power rather than of kind.

What no cut of this application changes is the fee. With the charge applied to both sides of the switch (and the premium turns over *less* than HAR, 4.57 against 5.02, so costs run in its favor), the unconstrained switch still pays Δ_γ of -0.14% to -0.78% per year: the risk reduction is bought with a lower mean return, and a mean-variance investor does not pay for it. The market-neutral panel turns marginally positive (0.35% and 0.19%), the one qualification this result carries, and we read it as the constraint protecting incoherent competitors rather than as an economic-value claim. Around the main cells, the grid says the same thing more loudly. A covariance matrix built on the variance risk premium is actively destructive (realized volatility of 17.19% to 17.6% across rungs; $\gamma = 10$ fees of -21.94% to -33.93%). The implied betas are the one cell that pays: Δ_γ of 1.54% and 1.75% on HAR, the table's best Sharpe ratio (0.91 against 0.76), surviving the constraint (1.44% and 1.72%). But the fee does not extend beyond HAR (-1.92% at $\gamma = 1$ on SHAR-F), and one paying cell among 26 models, sitting on the same block whose forecasting residual at SHAR-F-L is suggestive, unestablished, and, unlike the premium's gain, not attributable to physical history (Section 4.1), is an open question rather than a claim.

The two portfolio studies behind the Introduction's split record bear directly on this table. DeMiguel et al. (2013) and Kempf et al. (2015) test option-implied correlations by swapping them into an otherwise historical covariance matrix, and in both constructions the option market enters the pair structure as a level: a single monthly scalar shifting every historical correlation in the first, a single implied correlation common to every pair in the second. Neither swap improves the portfolios significantly; the value those papers do find comes through implied variances and, in the first, through implied skewness read as an expected-return signal, never through the correlations. That correlation failure is attributed to instability, a matrix that moves so much month to month that its weights churn without compensation. That diagnosis names from the outside the mechanism the Non-PD column measures from within: a market-wide level forced into a correlation matrix strains exactly the property a minimum-variance portfolio consumes. The grid reproduces the failure with its own materials, the implied level excluded at the design stage (Section 2.3) and the history-free premium, the grid's nearest counterpart to the object those papers swap in, breaking every matrix it enters.

Where they exchange the matrix, this application feeds the forecast, so the same option-market correlation view meets the portfolio in two forms, and the table can separate what the record could not: the failure belongs to the level's form, since anchored on the pair's realized pattern the same content is what repairs. The verdicts still meet at one boundary, no significant improvement there and no fee here, so the disagreement with the record is over attribution rather than outcome. On the variance channel, where both papers find their gains (in the case of Kempf et al. (2015), concentrated in crisis months of a kind our post-crisis window does not contain), this application is silent by construction: the volatility forecast of Equation (22) is common to every cell precisely so that the table reads the correlation model alone.

5.2. Forecasting the equity premium

Pollet and Wilson (2010) show that the average pairwise correlation among stocks manifests aggregate, undiversifiable risk and should predict the market's excess return; a more accurate forecast of expected average correlation should then predict it more sharply, which is the economic test this margin puts to the grid. For each model θ we collapse its pairwise forecasts into the equal-weighted average predicted correlation over the common pair set,

$$\overline{RC}_t^{\theta} = \frac{1}{|\mathcal{O}_t|} \sum_{(i,j)\in\mathcal{O}_t} \widehat{RC}_{ij,t+1}^{\theta}, \tag{26}$$

known at the end of month t; the historical benchmark is the lagged realized correlation RC^m averaged over the same pairs, the measure Pollet and Wilson (2010) themselves test. Alongside these we take the market correlation risk premium $\overline{CRP}_t$ as a predictor in its own right: the one purely forward-looking, model-free reading this design points at; per Section 2.3, it is also the one place the premium form is a genuine forward-looking claim rather than a reparametrization. Each measure enters the predictive regression

$$r_{t+1}^{M} = a + b\, g_t + c' z_t + \varepsilon_{t+1}, \qquad g_t \in \left\{\overline{RC}_t^{\theta},\, \overline{CRP}_t\right\}, \tag{27}$$

with r_{t+1}^{M} the CRSP value-weighted excess return and z_t, in the augmented form, the eight macroeconomic predictors of Welch and Goyal (2008); inference uses Newey–West standard errors. Out of sample, each measure forecasts r_{t+1}^{M} from an expanding univariate window and is scored against the prevailing mean (Campbell and Thompson, 2008; Dong, Li, Rapach and Zhou, 2022),

$$R^{2,\mathrm{EP}}_{\mathrm{OOS}} = 1 - \frac{\sum_t \left(r_{t+1}^{M} - \hat{r}_{t+1|t}^{M}\right)^2}{\sum_t \left(r_{t+1}^{M} - \bar{r}_t^{M}\right)^2}, \tag{28}$$

with significance from the DM test on squared errors. One design constraint matters here: our average-correlation series cannot begin until the rolling estimation window is consumed (first forecastable response month January 2010), so the conventional 10-year burn-in spends 71% of the series and leaves 48 months, 2020–2023, to score. We keep it as the baseline because it is the convention, and report a 5-year window (108 months) as robustness against the burn-in rather than as a competing headline.

Tables 6 and 7 report a null with clean edges. In sample, every average-correlation measure enters with a positive slope in the predicted direction, the forecast-based measures clustering tightly (t-statistics from 1.92 to 2.51) above the historical average correlation ($t = 1.69$). The forward-looking information adds nothing to that: the main specification's

Table 4
Global minimum-variance portfolios. Realized performance of the GMV portfolios (Equation (23)) formed from the covariance forecast (Equation (22)) of every model of Section 2.4: the four realized rungs, the twenty implied-augmented models, and the two penalized specifications, with HAR+CRP the main specification, each sharing the same HAR volatility forecast and convexity-corrected to a condition number below 10,000, over response months 2010–2023. All returns, Sharpe ratios, and fees are net of the ten-basis-point one-way charge on each portfolio's realized turnover. Columns report the annualized realized volatility σ (the mean of $\sqrt{w' RCov_{t+1} w}$, the primary criterion); the modified Diebold–Mariano statistic and its p-value on the paired monthly realized-variance differentials of each model against its *own* baseline, with a positive statistic favouring the model (Newey–West with 3 lags and the small-sample correction of Section 2.6; blank for the HAR benchmark, which has no baseline), where the baseline is the rung each augmented model extends, HAR for SHAR, SHAR-F, SHAR-F-L, and LASSO (realized), and LASSO (realized) for LASSO (realized+implied); "<Base", the share of months the model's realized variance sits below its baseline's; the annualized mean net return; the Sharpe ratio; the Fleming et al. (2003) fee Δ_γ (annualized, in %) for switching from the HAR covariance at risk aversion $\gamma \in \{1, 10\}$, on net returns; the average monthly turnover; and "Non-PD", the percentage of months in which the model's *raw* assembled correlation matrix, before the convexity correction, had a negative eigenvalue. The correction is applied identically to every specification, so Non-PD measures the forecast that entered it, not the matrix that left.

Covariance	σ (%)	DM	p	<Base (%)	Return (%)	Sharpe	$\Delta_{\gamma=1}$	$\Delta_{\gamma=10}$	Turnover	Non-PD (%)
Panel A: HAR										
HAR	9.55				7.24	0.76	-0.00	0.00	5.02	0.0
HAR+CRP	9.33	0.99	0.325	82.7	7.17	0.77	-0.14	-0.78	4.57	3.0
HAR+CRP-MW	15.68	-4.71	0.000	29.8	5.52	0.35	-3.31	-20.04	10.57	31.0
HAR+VRP	17.19	-2.50	0.014	35.7	1.42	0.08	-8.14	-33.93	9.55	44.6
HAR+skew	9.51	1.21	0.229	67.9	7.21	0.76	-0.03	-0.04	5.02	0.0
HAR+$\beta^{\mathbb{Q}}$	9.65	-1.56	0.120	36.3	8.76	0.91	1.54	1.75	5.07	0.0
Panel B: SHAR										
SHAR	11.21	-4.50	0.000	29.8	4.87	0.43	-2.49	-3.64	7.56	14.9
SHAR+CRP	10.30	3.76	0.000	84.5	5.13	0.50	-2.21	-3.24	6.14	11.9
SHAR+CRP-MW	15.74	-4.33	0.000	31.5	3.14	0.20	-5.56	-20.80	11.06	36.9
SHAR+VRP	17.44	-2.34	0.021	41.7	0.46	0.03	-8.89	-31.92	10.29	50.6
SHAR+skew	11.17	1.14	0.258	62.5	4.81	0.43	-2.56	-3.81	7.56	15.5
SHAR+$\beta^{\mathbb{Q}}$	11.37	-1.91	0.057	38.7	6.30	0.55	-1.07	-2.28	7.66	13.1
Panel C: SHAR-F										
SHAR-F	13.56	-4.72	0.000	28.0	4.49	0.33	-3.26	-8.34	10.04	38.7
SHAR-F+CRP	12.02	2.71	0.007	83.9	5.04	0.42	-2.48	-5.21	8.09	25.0
SHAR-F+CRP-MW	18.92	-4.97	0.000	32.1	2.25	0.12	-7.06	-29.52	13.84	55.4
SHAR-F+VRP	17.50	-2.46	0.015	50.0	2.15	0.12	-6.56	-21.94	10.73	60.1
SHAR-F+skew	13.87	-1.27	0.206	50.6	2.91	0.21	-4.96	-11.16	10.52	41.7
SHAR-F+$\beta^{\mathbb{Q}}$	13.46	0.82	0.413	44.6	5.80	0.43	-1.92	-6.66	9.93	34.5
Panel D: SHAR-F-L										
SHAR-F-L	12.92	-5.05	0.000	28.0	3.18	0.25	-4.31	-6.76	10.49	45.2
SHAR-F-L+CRP	12.58	0.90	0.371	72.6	2.88	0.23	-4.64	-7.34	9.78	38.7
SHAR-F-L+CRP-MW	15.56	-3.74	0.000	46.4	0.47	0.03	-7.81	-18.45	12.55	52.4
SHAR-F-L+VRP	17.60	-2.52	0.013	42.3	2.49	0.14	-6.28	-22.43	11.38	66.1
SHAR-F-L+skew	13.34	-1.91	0.058	54.2	1.35	0.10	-6.20	-9.21	11.04	49.4
SHAR-F-L+$\beta^{\mathbb{Q}}$	13.52	-3.47	0.001	22.6	3.27	0.24	-4.35	-8.13	11.02	44.6
Panel E: LASSO										
LASSO (realized)	9.61	-1.16	0.248	64.3	7.64	0.79	0.46	1.03	5.52	5.4
LASSO (realized+implied)	16.94	-3.12	0.002	25.6	-1.22	-0.07	-10.30	-30.08	10.46	42.3

slope ($t = 1.92$) is no larger and no sharper than HAR's ($t = 2.01$), the same holds for every premium-augmented model against its rung, and the market correlation risk premium $\overline{CRP}$, the purest forward-looking measure in the design, does not predict the premium at all ($t = 0.31$). Once the Welch–Goyal predictors enter, every correlation-based measure collapses ($|t|$ below 1.4 throughout); the single cell above two, a VRP-augmented row at $|t| = 2.41$, is the extreme order statistic of a 26-row table and we read it as such. Out of sample the story repeats with a caution attached. The main specification posts 2.28% ($p = 0.416$), insignificant like everything else at the baseline burn-in; $\overline{CRP}$ is *significantly worse* than the prevailing mean on 48 months ($R^{2,\mathrm{EP}}_{\mathrm{OOS}} = -1.09\%$, $p = 0.036$) and merely unremarkable

Table 5
Global minimum-variance portfolios, market-neutral. As Table 4 (same forecasts, same universe, same convexity correction, same net-of-cost convention, same columns), but the weights additionally satisfy $w'\beta_t = 0$ (Equation (25)), with β_t the sixty-month rolling market betas of Section 2.2. The Diebold–Mariano statistic tests each model's realized portfolio variance against its own baseline's *market-neutral* portfolio, so the constraint's effect on risk, common to every model, is differenced out and the statistic reads the correlation forecast alone; the fee likewise compares each model with HAR under the same constraint. Non-PD describes the forecast rather than the weights and is therefore identical to Table 4's, repeated for readability.

Covariance	σ (%)	DM	p	<Base (%)	Return (%)	Sharpe	$\Delta_{\gamma=1}$	$\Delta_{\gamma=10}$	Turnover	Non-PD (%)
Panel A: HAR										
HAR	9.66				7.07	0.73	-0.00	-0.00	5.04	0.0
HAR+CRP	9.33	2.33	0.021	84.5	7.43	0.80	0.35	0.19	4.44	3.0
HAR+CRP-MW	14.13	-3.99	0.000	33.3	7.86	0.56	-0.21	-10.40	9.10	31.0
HAR+VRP	14.26	-3.26	0.001	35.1	2.05	0.14	-5.58	-11.15	8.07	44.6
HAR+skew	9.62	1.27	0.206	70.2	7.07	0.73	0.01	-0.00	5.03	0.0
HAR+$\beta^{\mathbb{Q}}$	9.75	-1.40	0.163	34.5	8.47	0.87	1.44	1.72	5.09	0.0
Panel B: SHAR										
SHAR	11.12	-4.26	0.000	29.8	4.84	0.44	-2.34	-3.45	7.30	14.9
SHAR+CRP	10.18	4.30	0.000	88.1	5.92	0.58	-1.20	-1.69	5.82	11.9
SHAR+CRP-MW	14.10	-3.52	0.001	35.1	6.43	0.46	-1.52	-10.36	9.42	36.9
SHAR+VRP	14.67	-2.76	0.007	42.9	0.93	0.06	-6.71	-12.41	8.86	50.6
SHAR+skew	11.08	0.93	0.352	64.9	4.80	0.43	-2.39	-3.60	7.31	15.5
SHAR+$\beta^{\mathbb{Q}}$	11.23	-1.65	0.101	39.3	6.04	0.54	-1.13	-2.10	7.39	13.1
Panel C: SHAR-F										
SHAR-F	11.86	-4.86	0.000	30.4	5.25	0.44	-2.08	-4.64	8.23	38.7
SHAR-F+CRP	10.71	3.78	0.000	83.9	6.21	0.58	-0.93	-1.67	6.41	25.0
SHAR-F+CRP-MW	14.47	-4.10	0.000	35.1	5.92	0.41	-2.07	-11.40	9.88	55.4
SHAR-F+VRP	14.49	-2.30	0.023	42.3	3.37	0.23	-4.17	-8.81	8.59	60.1
SHAR-F+skew	12.12	-1.73	0.086	50.0	4.12	0.34	-3.27	-6.49	8.65	41.7
SHAR-F+$\beta^{\mathbb{Q}}$	11.89	-0.44	0.661	38.7	7.56	0.64	0.26	-2.00	8.13	34.5
Panel D: SHAR-F-L										
SHAR-F-L	11.25	-4.32	0.000	42.9	7.36	0.65	0.30	0.35	8.26	45.2
SHAR-F-L+CRP	10.80	3.59	0.000	78.6	7.83	0.73	0.79	0.95	7.41	38.7
SHAR-F-L+CRP-MW	13.12	-3.77	0.000	47.6	4.69	0.36	-2.86	-7.54	9.80	52.4
SHAR-F-L+VRP	13.70	-2.21	0.028	39.9	6.10	0.45	-1.33	-4.90	8.31	66.1
SHAR-F-L+skew	11.59	-1.69	0.094	54.2	6.11	0.53	-0.98	-1.16	8.75	49.4
SHAR-F-L+$\beta^{\mathbb{Q}}$	11.59	-2.53	0.012	23.2	7.89	0.68	0.78	0.35	8.60	44.6
Panel E: LASSO										
LASSO (realized)	9.37	1.88	0.061	72.0	7.76	0.83	0.79	1.74	5.08	5.4
LASSO (realized+implied)	14.22	-2.78	0.006	28.0	0.91	0.06	-6.80	-13.11	8.45	42.3

on 108 ($p = 0.206$). A result that reverses when the window doubles is an artifact of the burn-in, and we decline to report either as a finding. The penalized LASSO over both blocks posts the table's largest in-sample t (2.36) and largest out-of-sample R^2 (5.39%, $p = 0.158$); it collapses under the controls ($t = 1.34$), which is what one expects of the extreme order statistic of a grid this size.

The shape of the null is the informative part. A pairwise forecast's content, here mostly which pairs' correlations persist, is a statement about the *cross-section* of comovement, and Equation (26) averages that away; what remains is a common series the realized history already tracks. The behavior of $\overline{CRP}$ makes the point without inference: the aggregate premium is the design's one genuinely forward-looking claim, and it carries nothing about the market's return on this sample. Two boundaries travel with that reading. Our window begins after the rolling window is consumed and so excludes the 2008 crisis, where out-of-sample equity-premium predictability concentrates (Rapach et al., 2010); and Buss et al. (2019) find that expected correlation extracted jointly from index and single-name options does predict returns at longer horizons over a crisis-inclusive sample; their measure, their horizons, and their sample all differ, and the two results are consistent. At the monthly horizon, on a post-crisis sample, better pairwise correlation forecasts do not become sharper aggregate return prediction.

Table 6
Equity-premium prediction: in-sample regressions. Predictive regressions of the one-month-ahead CRSP value-weighted excess return on each average-correlation measure (Equation (27)), estimated over response months 2010–2023. Each forecast-based measure is the equal-weighted average predicted correlation over the common evaluation sample $\mathcal{O}$ (Equation (26)), reported for the historical benchmark RC^m and for every model of the grid of Section 2.4; the market correlation risk premium $\overline{CRP}$ enters as a model-free predictor. HAR+CRP is the main specification; the remaining models are a robustness grid. "Simple" regressions include the measure alone; "with controls" add the eight Welch–Goyal predictors. t-statistics (in parentheses) use Newey–West standard errors.

	Simple			With controls			
Predictor	b	t	Adj. R^2	b	t	Adj. R^2	N
Historical	0.046	(1.69)	0.94	-0.005	(-0.12)	4.83	168
Panel A: HAR							
HAR	0.099	(2.01)	1.49	0.023	(0.28)	4.88	168
HAR+CRP	0.111	(1.92)	2.33	0.044	(0.37)	5.03	168
HAR+CRP-MW	0.128	(2.05)	2.47	0.053	(0.46)	5.09	168
HAR+VRP	0.126	(2.45)	3.79	0.074	(1.02)	5.34	168
HAR+skew	0.096	(2.00)	1.44	0.023	(0.27)	4.88	168
HAR+$\beta^{\mathbb{Q}}$	0.101	(2.04)	1.58	0.031	(0.37)	4.93	168
Panel B: SHAR							
SHAR	0.101	(2.04)	1.60	0.027	(0.33)	4.90	168
SHAR+CRP	0.112	(1.95)	2.39	0.046	(0.41)	5.06	168
SHAR+CRP-MW	0.130	(2.09)	2.59	0.057	(0.51)	5.14	168
SHAR+VRP	0.127	(2.43)	3.74	0.071	(0.98)	5.33	168
SHAR+skew	0.099	(2.03)	1.57	0.027	(0.33)	4.90	168
SHAR+$\beta^{\mathbb{Q}}$	0.103	(2.06)	1.68	0.035	(0.41)	4.95	168
Panel C: SHAR-F							
SHAR-F	0.102	(2.10)	1.81	0.033	(0.41)	4.95	168
SHAR-F+CRP	0.110	(1.93)	2.41	0.044	(0.40)	5.05	168
SHAR-F+CRP-MW	0.125	(2.07)	2.62	0.055	(0.51)	5.15	168
SHAR-F+VRP	0.126	(2.50)	4.00	0.079	(1.12)	5.48	168
SHAR-F+skew	0.099	(2.08)	1.73	0.033	(0.39)	4.95	168
SHAR-F+$\beta^{\mathbb{Q}}$	0.103	(2.10)	1.84	0.039	(0.47)	5.00	168
Panel D: SHAR-F-L							
SHAR-F-L	0.109	(2.43)	2.29	0.113	(1.35)	6.05	168
SHAR-F-L+CRP	0.113	(2.38)	2.57	0.095	(1.08)	5.77	168
SHAR-F-L+CRP-MW	0.114	(2.51)	2.50	0.107	(1.22)	5.97	168
SHAR-F-L+VRP	0.136	(2.50)	4.56	0.180	(2.41)	7.38	168
SHAR-F-L+skew	0.107	(2.41)	2.18	0.114	(1.37)	6.04	168
SHAR-F-L+$\beta^{\mathbb{Q}}$	0.109	(2.42)	2.26	0.111	(1.33)	6.01	168
Panel E: LASSO							
LASSO (realized)	0.106	(2.28)	1.65	0.080	(0.94)	5.37	168
LASSO (realized+implied)	0.137	(2.36)	3.58	0.110	(1.34)	5.79	168
Corr. risk premium $\overline{CRP}$	0.008	(0.31)	-0.54	0.024	(0.98)	5.22	168

5.3. Pairs trading

Pairs trading reads the forecasts as a cross-sectional ranking and never aggregates or inverts them: the margin built to suit a pairwise signal, following Chen, Chen, Chen and Li (2019) and Gatev, Goetzmann and Rouwenhorst (2006) as implemented by Bollerslev et al. (2026). For stock i in month t, its pair portfolio $\mathcal{P}_{i,t}$ is its twenty most-correlated

Table 7
Equity-premium prediction: out-of-sample evaluation. Out-of-sample R^2 of each average-correlation measure relative to the prevailing-mean benchmark (Equation (28)), from expanding-window univariate forecasts of the one-month-ahead market excess return, with modified Diebold–Mariano statistics and p-values. Measures are those of Table 6. The baseline primes the expanding window with 10 years of data, leaving an out-of-sample period of 2020–2023(48 months); the 5-year-window columns, whose out-of-sample period runs 2015–2023(108 months), are reported as robustness against the burn-in and are not a competing headline.

	10-year burn-in (baseline)			5-year burn-in		
Predictor	R^2_{OOS} (%)	DM stat	DM p	R^2_{OOS} (%)	DM stat	DM p
Historical	0.83	0.45	0.658	0.78	0.56	0.577
Panel A: HAR						
HAR	1.44	0.80	0.429	1.51	0.94	0.348
HAR+CRP	2.28	0.82	0.416	2.50	1.12	0.263
HAR+CRP-MW	2.58	0.87	0.388	2.67	1.12	0.264
HAR+VRP	5.42	1.18	0.243	4.04	1.22	0.224
HAR+skew	1.37	0.78	0.441	1.51	0.96	0.340
HAR+$\beta^{\mathbb{Q}}$	1.65	0.94	0.350	1.61	0.99	0.322
Panel B: SHAR						
SHAR	1.60	0.86	0.392	1.56	0.94	0.349
SHAR+CRP	2.47	0.92	0.361	2.57	1.18	0.240
SHAR+CRP-MW	2.81	0.96	0.340	2.73	1.15	0.253
SHAR+VRP	5.31	1.22	0.227	3.92	1.24	0.218
SHAR+skew	1.53	0.84	0.403	1.57	0.96	0.340
SHAR+$\beta^{\mathbb{Q}}$	1.79	1.00	0.321	1.63	0.98	0.327
Panel C: SHAR-F						
SHAR-F	1.78	0.93	0.358	1.85	1.06	0.293
SHAR-F+CRP	2.32	0.82	0.418	2.59	1.13	0.261
SHAR-F+CRP-MW	2.64	0.86	0.393	2.77	1.12	0.267
SHAR-F+VRP	5.67	1.27	0.211	4.33	1.32	0.190
SHAR-F+skew	1.66	0.89	0.377	1.81	1.06	0.293
SHAR-F+$\beta^{\mathbb{Q}}$	1.90	1.04	0.306	1.86	1.07	0.286
Panel D: SHAR-F-L						
SHAR-F-L	2.41	1.26	0.214	2.55	1.34	0.182
SHAR-F-L+CRP	3.15	1.75	0.087	3.24	1.85	0.067
SHAR-F-L+CRP-MW	3.15	1.90	0.064	3.13	1.86	0.066
SHAR-F-L+VRP	6.72	1.28	0.207	5.34	1.39	0.167
SHAR-F-L+skew	2.29	1.26	0.215	2.46	1.34	0.185
SHAR-F-L+$\beta^{\mathbb{Q}}$	2.38	1.30	0.200	2.47	1.30	0.195
Panel E: LASSO						
LASSO (realized)	1.62	0.88	0.386	1.51	0.81	0.422
LASSO (realized+implied)	5.39	1.44	0.158	4.17	1.48	0.142
Corr. risk premium $\overline{CRP}$	-1.09	-2.16	0.036	-1.48	-1.27	0.206

peers by the twelve-month history

$$RC^{\text{hist}}_{ij,t} = \frac{1}{12}\sum_{s=1}^{12} RC_{ij,t-s}, \tag{29}$$

drawn from the stocks of $\mathcal{O}$ and common to every model. The trading signal is the return divergence of Chen et al. (2019),

$$RetDiff_{i,t} = \beta_{i,t}\left(PRet_{i,t} - r_{f,t}\right) - \left(Ret_{i,t} - r_{f,t}\right), \tag{30}$$

with $PRet_{i,t}$ the pair portfolio's return and $\beta_{i,t}$ from regressing i's daily returns on its pair portfolio's over the trailing year: a stock that underperformed its peers should outperform if the pair reconverges. Convergence presumes the pair still comoves, which is where a forecast enters: each model nominates the stocks whose predicted comovement with their peers rises most. We condition on the *rank* difference

$$\Delta RC^{\theta}_{i,t} = F_t\big(\widehat{RC}^{\theta}_{i,t+1}\big) - F_t\big(RC^{\text{hist}}_{i,t}\big), \tag{31}$$

F_t the month-t empirical distribution function, rather than on the level difference of Bollerslev et al. (2026): the models differ in how far they shrink their forecasts (cross-sectional dispersion between 0.53 and 0.67 times the realized correlation's), and a level difference rewards dispersion for reasons unrelated to accuracy, where ranks are invariant to any monotone rescaling. Stocks sort into $G = 5$ quintiles on ΔRC^{θ}; within the top quintile, a long–short spread on $RetDiff$, held one month, equal- and value-weighted. The criterion is the four-factor alpha (Fama and French, 1993; Carhart, 1997) of the spread, not its raw mean: the conditioning variable is a risk premium, and a spread inside a subsample selected on priced correlation risk could be compensation for that risk. Beside the net-of-cost columns we report the break-even one-way cost at which the alpha vanishes, which is affine in the charge,

$$c^{*} = \alpha \,/\, \bar{\tau}, \qquad \bar{\tau} = \text{the intercept from projecting turnover on the same four factors.} \tag{32}$$

Table 8 reports a null with the right ordering. The unconditional $RetDiff$ spread, conditioning on no forecast, earns 0.47% a year ($t = 0.34$): return divergence applied blind is worthless. Conditioning on a correlation forecast moves every point estimate the way the forecasting results predict, and adding CRP raises the alpha's t-statistic at every rung (from 0.31 to 1.27 on HAR, 0.04 to 1.21 on SHAR, 0.41 to 1.47 on SHAR-F), with the main specification earning 2.91% ($t = 1.27$; value-weighted 4.04%, $t = 1.14$, Table 9). The ordering is where the good news ends: no cell in the grid reaches significance (the largest $|t|$ anywhere is 1.92), and the power arithmetic says the window could not have done better. At these spreads' volatility, the smallest alpha the sample could have called significant is 4.57% for the main cell (4.57%–6.55% across Table 8). That places our null squarely on the HAR-conditioned spread of Bollerslev et al. (2026) (3.63% per year, $t = 0.88$), not short of it; their one significant pairs alpha belongs to their penalized specification, whose implied block is a forecasting null in our Table 2. Costs close whatever remains. Turnover is structural (the conditioning quintile is redrawn monthly, 3.33 to 3.77 per month across the grid), so the ten-basis-point charge takes the main cell's alpha to −1.38% ($t = -0.6$), and the break-even cost of Equation (32) is 6.8 basis points one-way (−3.9 to 9.6 across Table 8). That break-even is of the same order as the effective spread a large-capitalization book actually pays, so the net column does not say the strategy is uneconomic; what c^{*} prices, either way, is a gross alpha indistinguishable from zero.

Two untabulated variants sharpen the reading without changing it. Under the level rather than the rank definition of ΔRC the premium's alphas are larger (4.33%, $t = 1.55$, for the main cell), the direction the dispersion argument behind Equation (31) predicts, and we read the gap as the artifact the rank definition removes; no cell is significant under either definition. And a buffered conditioning sort (Novy-Marx and Velikov, 2016), in which a stock enters the eligible set in the top $1/G$ and remains while in the top q, raises most point estimates and produces the sweep's only significant cells: at $q = 0.4$ the main specification reaches 4.62% with $t = 2.54$, the largest statistic among the 52 buffered cells, of which 2 clear two. We report it and do not promote it: the profile is non-monotone in q (the same cell falls back to $t = 1.67$ at $q = 0.6$), which is the signature of noise rather than persistence, and net of costs that cell retains 0.52%. Every reading of the variants leaves the conclusion where the unbuffered table put it.

5.4. Risk targeting

The fourth margin asks the forecast only to *measure*. A manager holds a portfolio chosen on other grounds and asks the covariance forecast how risky next month will be: nothing is optimized, nothing inverted, and no t=2 hurdle against return noise intervenes. Of the four margins, this is the one where accuracy converts most directly. The test portfolios are 14 long-short quintile strategies built from the anomaly characteristics already ingested for the factor-driven features (Section 2.2), equal-weighted within legs, weights common to every model. Each model prices each position through its covariance forecast (Equation (22), shared volatilities, identical convexity correction[5]), and calibration is the ratio

[5] Bollerslev et al. (2026) state no correction for this application; our variance-risk-premium models assemble a non-positive-definite matrix in roughly half the months (Table 4), and a negative forecast *variance* is not a reportable risk number. The correction is the one applied everywhere else, identically across models.

Table 8
Pairs trading, equal-weighted. Performance of the double-sorted long–short pairs strategy (Section 5.3) for every model of Section 2.4, over the 168 holding months, 2010–2023. Each stock's pair portfolio is its twenty most-correlated peers by twelve-month realized-correlation history; stocks are sorted into quintiles on the predicted correlation change ΔRC^{θ} (Equation (31), the rank definition) and, within the top quintile, long–short on return divergence $RetDiff$ (Equation (30)), the spread equal-weighted. "Unconditional" sorts on $RetDiff$ over the whole cross-section, conditioning on no forecast, and sits outside the family panels. Columns report the annualized four-factor (Fama and French, 1993; Carhart, 1997) alpha α and its Newey–West t-statistic (the criterion fixed in Section 5.3), the annualized raw mean spread, its net-of-cost counterpart at the ten-basis-point one-way charge, the net-of-cost four-factor alpha and its t, and the average monthly turnover.

Model	α (%)	$t(\alpha)$	Raw (%)	Net (%)	Net α	t	c^* (bp)	Turnover
Unconditional	0.47	0.34	2.39	–	–	–	–	–
Panel A: HAR								
HAR	0.83	0.31	3.25	-1.13	-3.54	-1.33	1.9	3.66
HAR+CRP	2.91	1.27	4.27	-0.03	-1.38	-0.60	6.8	3.61
HAR+CRP-MW	0.86	0.33	2.44	-1.93	-3.51	-1.37	2.0	3.67
HAR+VRP	3.22	1.12	5.43	1.07	-1.13	-0.39	7.4	3.66
HAR+skew	1.88	0.78	4.43	0.05	-2.49	-1.03	4.3	3.67
HAR+$\beta^{\mathbb{Q}}$	2.98	1.03	4.94	0.60	-1.36	-0.47	6.9	3.64
Panel B: SHAR								
SHAR	0.11	0.04	2.28	-2.09	-4.26	-1.49	0.3	3.67
SHAR+CRP	3.07	1.21	4.39	0.09	-1.22	-0.48	7.2	3.60
SHAR+CRP-MW	-0.51	-0.17	1.06	-3.32	-4.89	-1.61	–	3.68
SHAR+VRP	4.16	1.50	5.77	1.41	-0.19	-0.07	9.6	3.65
SHAR+skew	2.82	1.04	5.24	0.86	-1.55	-0.57	6.5	3.67
SHAR+$\beta^{\mathbb{Q}}$	2.41	0.76	4.14	-0.21	-1.94	-0.61	5.5	3.65
Panel C: SHAR-F								
SHAR-F	1.29	0.41	3.69	-0.65	-3.03	-0.97	3.0	3.64
SHAR-F+CRP	3.46	1.47	4.78	0.47	-0.84	-0.36	8.0	3.61
SHAR-F+CRP-MW	0.26	0.09	2.17	-2.18	-4.08	-1.49	0.6	3.64
SHAR-F+VRP	3.00	1.13	4.96	0.62	-1.33	-0.50	6.9	3.64
SHAR-F+skew	2.30	0.79	4.44	0.09	-2.04	-0.70	5.3	3.65
SHAR-F+$\beta^{\mathbb{Q}}$	2.30	0.70	3.73	-0.61	-2.03	-0.62	5.3	3.64
Panel D: SHAR-F-L								
SHAR-F-L	0.04	0.02	1.51	-2.79	-4.26	-1.68	0.1	3.60
SHAR-F-L+CRP	2.67	1.07	4.14	-0.14	-1.59	-0.64	6.3	3.59
SHAR-F-L+CRP-MW	-1.66	-0.72	-0.21	-4.51	-5.96	-2.57	–	3.61
SHAR-F-L+VRP	0.27	0.12	1.55	-2.75	-4.03	-1.74	0.6	3.61
SHAR-F-L+skew	-1.42	-0.47	0.70	-3.59	-5.70	-1.89	–	3.60
SHAR-F-L+$\beta^{\mathbb{Q}}$	-1.00	-0.41	0.65	-3.63	-5.28	-2.15	–	3.59
Panel E: LASSO								
LASSO (realized)	-0.93	-0.31	1.07	-3.19	-5.19	-1.71	–	3.57
LASSO (realized+implied)	-0.13	-0.05	1.63	-2.64	-4.40	-1.83	–	3.58

of forecast to realized portfolio variance (Bollerslev et al., 2026, Eq. 11),

$$\text{AvgRatio}^{\theta} = \frac{1}{T}\sum_{t=1}^{T} \frac{w_t' \,\widehat{RCov}^{\theta}_{t+1}\, w_t}{w_t' \, RCov_{t+1}\, w_t}, \tag{33}$$

Table 9
Pairs trading, value-weighted. As Table 8, with the spread value-weighted by market capitalization.

Model	α (%)	$t(\alpha)$	Raw (%)	Net (%)	Net α	t	c^* (bp)	Turnover
Unconditional	1.38	0.58	2.37	–	–	–	–	–
Panel A: HAR								
HAR	-1.04	-0.30	2.64	-1.84	-5.51	-1.59	–	3.75
HAR+CRP	4.04	1.14	5.65	1.20	-0.38	-0.11	9.1	3.73
HAR+CRP-MW	-2.22	-0.71	0.83	-3.62	-6.66	-2.13	–	3.73
HAR+VRP	1.46	0.42	4.66	0.20	-2.98	-0.86	3.3	3.74
HAR+skew	0.85	0.29	4.39	-0.09	-3.61	-1.21	1.9	3.75
HAR+$\beta^{\mathbb{Q}}$	3.19	0.83	5.99	1.54	-1.24	-0.33	7.2	3.73
Panel B: SHAR								
SHAR	2.91	0.81	5.72	1.24	-1.55	-0.43	6.5	3.76
SHAR+CRP	4.46	1.19	5.90	1.48	0.07	0.02	10.2	3.71
SHAR+CRP-MW	-1.10	-0.34	2.00	-2.47	-5.56	-1.72	–	3.75
SHAR+VRP	3.71	1.11	6.64	2.18	-0.73	-0.22	8.4	3.73
SHAR+skew	3.09	1.02	6.89	2.40	-1.39	-0.46	6.9	3.77
SHAR+$\beta^{\mathbb{Q}}$	1.95	0.46	5.03	0.58	-2.49	-0.59	4.4	3.73
Panel C: SHAR-F								
SHAR-F	-0.15	-0.04	2.84	-1.59	-4.56	-1.24	–	3.72
SHAR-F+CRP	0.02	0.01	1.90	-2.51	-4.37	-1.34	0.0	3.70
SHAR-F+CRP-MW	-1.99	-0.57	0.86	-3.56	-6.38	-1.84	–	3.70
SHAR-F+VRP	0.02	0.00	2.21	-2.21	-4.38	-1.23	0.0	3.70
SHAR-F+skew	-0.25	-0.07	2.11	-2.32	-4.66	-1.28	–	3.72
SHAR-F+$\beta^{\mathbb{Q}}$	0.97	0.23	3.46	-1.00	-3.48	-0.84	2.2	3.74
Panel D: SHAR-F-L								
SHAR-F-L	2.03	0.62	2.81	-1.63	-2.38	-0.73	4.6	3.72
SHAR-F-L+CRP	2.97	0.92	3.49	-0.93	-1.41	-0.44	6.8	3.71
SHAR-F-L+CRP-MW	-5.47	-1.68	-4.12	-8.53	-9.88	-3.02	–	3.70
SHAR-F-L+VRP	2.11	0.56	2.42	-2.04	-2.33	-0.62	4.7	3.74
SHAR-F-L+skew	-1.21	-0.30	0.57	-3.83	-5.59	-1.39	–	3.69
SHAR-F-L+$\beta^{\mathbb{Q}}$	0.36	0.10	2.01	-2.42	-4.05	-1.10	0.8	3.71
Panel E: LASSO								
LASSO (realized)	-2.48	-0.72	-0.63	-5.01	-6.85	-1.98	–	3.68
LASSO (realized+implied)	-1.18	-0.33	0.24	-4.17	-5.58	-1.55	–	3.69

with unity the target and a bootstrapped 95% confidence interval per strategy.[6] The same volatility forecasts imply a 95% monthly value-at-risk, backtested on strategy returns net of the turnover charge with the unconditional-coverage test of Kupiec (1995) and the conditional-coverage test of Christoffersen (1998); the ratio itself involves no trading and carries no charge.

The answer in Table 10 is not close: every forecast in the grid materially understates anomaly-portfolio risk. Mean ratios run from 0.63 to 0.85 across the 26 models, and unity falls inside the bootstrap interval for not one strategy for any of them, where Bollerslev et al. (2026) report 0.83 for HAR and 1.02 for their LASSO. The VaR backtest says the same in loss space: violations in 7.9% to 11.1% of months against a five-percent target. On the main comparison the premium makes matters worse, not better (mean ratio 0.63 against HAR's 0.69, violations 10.9% against 10%, conditional coverage passing 5 of 14 strategies against 6): a small, consistent negative rather than a null, with the implied betas the mirror case (0.85, closest to unity and still far). The memo row locates the failure: pricing the same portfolios with the month's own *realized* correlations and the forecast volatilities yields a mean ratio of 0.96, unity inside the interval for 9 of 14 strategies, and violations of 6.8%. The shared volatility forecasts account for a few points

[6]Following Bollerslev et al. (2026, fn. 24): the monthly ratios are drawn with replacement, each draw is averaged, and the interval is the 2.5th–97.5th percentile of the bootstrap averages.

Table 10
Risk targeting. Calibration of each model's portfolio-risk forecasts on the 14 long-short quintile strategies (Section 5.4), over the 168 evaluation months. "Mean ratio" is AvgRatio (Equation (33)) averaged across strategies; $|\cdot - 1|$ its mean absolute distance from unity; "Unity in CI" the number of strategies whose bootstrapped 95% interval covers one. The remaining columns backtest the implied 95% monthly VaR on net-of-cost strategy returns: the mean violation rate (target 5%), and the number of strategies for which the Kupiec (1995) unconditional and Christoffersen (1998) conditional coverage tests do not reject at the 5% level. The memo row prices the same portfolios with the month's own realized correlations and the forecast volatilities, isolating the volatility channel every model shares; its distance from unity is the part of the shortfall no correlation forecast could remove.

Correlation forecast	Mean ratio	$\lvert\cdot - 1\rvert$	Unity in CI	Viol. (%)	Kupiec	Christ.
Panel A: HAR						
HAR	0.69	0.31	0/14	10.0	4/14	6/14
HAR+CRP	0.63	0.37	0/14	10.9	4/14	5/14
HAR+CRP-MW	0.68	0.32	0/14	10.2	5/14	6/14
HAR+VRP	0.64	0.36	0/14	11.0	3/14	6/14
HAR+skew	0.69	0.31	0/14	10.1	4/14	6/14
HAR+$\beta^{\mathbb{Q}}$	0.71	0.29	0/14	9.5	6/14	6/14
Panel B: SHAR						
SHAR	0.68	0.32	0/14	10.0	5/14	5/14
SHAR+CRP	0.63	0.37	0/14	11.1	5/14	5/14
SHAR+CRP-MW	0.68	0.32	0/14	10.4	5/14	5/14
SHAR+VRP	0.64	0.36	0/14	10.9	4/14	5/14
SHAR+skew	0.68	0.32	0/14	10.1	5/14	5/14
SHAR+$\beta^{\mathbb{Q}}$	0.71	0.29	0/14	9.5	5/14	6/14
Panel C: SHAR-F						
SHAR-F	0.72	0.28	0/14	9.2	5/14	7/14
SHAR-F+CRP	0.66	0.34	0/14	10.2	5/14	6/14
SHAR-F+CRP-MW	0.71	0.29	0/14	9.8	5/14	6/14
SHAR-F+VRP	0.68	0.32	0/14	10.2	5/14	6/14
SHAR-F+skew	0.71	0.29	0/14	9.3	5/14	7/14
SHAR-F+$\beta^{\mathbb{Q}}$	0.74	0.26	0/14	8.9	6/14	7/14
Panel D: SHAR-F-L						
SHAR-F-L	0.84	0.16	0/14	8.0	8/14	10/14
SHAR-F-L+CRP	0.79	0.21	0/14	8.5	7/14	8/14
SHAR-F-L+CRP-MW	0.82	0.18	0/14	8.4	8/14	9/14
SHAR-F-L+VRP	0.77	0.23	0/14	8.7	6/14	8/14
SHAR-F-L+skew	0.84	0.16	0/14	7.9	9/14	10/14
SHAR-F-L+$\beta^{\mathbb{Q}}$	0.85	0.15	0/14	7.9	9/14	10/14
Panel E: LASSO						
LASSO (realized)	0.84	0.16	0/14	8.0	7/14	8/14
LASSO (realized+implied)	0.75	0.25	0/14	9.4	5/14	6/14
Realized correlation (memo)	0.96	0.04	9/14	6.8	11/14	12/14

of the shortfall; the rest belongs to the correlation forecasts. The mechanism is the one from Section 5.1, with the opposite sign. A pooled forecast compresses the cross-section of correlations toward its conditional mean; a long-short spread's variance loads on within-leg minus cross-leg correlation, the very dispersion being compressed; and the premium-augmented models understate most because the premium adds *common* structure by construction. Common structure is what a covariance matrix lacks (the repair of Section 5.1) and what a bet on correlation dispersion cannot afford.

Taken together, the four applications draw one boundary from four directions, with every return-based comparison charged. Inverted into minimum-variance weights, the premium delivers the section's positive result, significant reductions in realized portfolio risk that a forecasting-equivalent placebo cannot reproduce, and still earns no fee,

because the risk reduction is bought with mean return. Collapsed to a scalar, it has nothing to say that realized history does not. Read as a cross-sectional ranking, it orders the grid as its accuracy predicts and earns nothing distinguishable from zero. Asked to price a dispersion bet, it is short with everything else in the grid, and shortest of all. The premium's value in portfolio choice is confined to objects made of coherence, and runs against objects made of dispersion; the Conclusion draws the two mechanisms together.

6. Robustness

Nothing in this section re-estimates a model. Both checks are read off the same fitted forecasts and the same evaluation sample $\mathcal{O}$ that produce the main tables; what varies is how the pair-months are partitioned, by calendar in Section 6.1 and by firm size in Section 6.2. The free parameters of the economic applications are swept where those applications are read: the buffered conditioning sort of the pairs strategy in Section 5.3, and the burn-in of the equity-premium window in Section 5.2.

6.1. Subsample stability

We partition the 168 response months into the three calendar blocks of Table 11, of comparable length and cut on calendar boundaries rather than on anything the results could suggest, and report each model's marginal contribution and DM p-value within each (the analogue of Bollerslev et al. (2026, Table 7)), together with the across-block test of Equation (20), which spends all 168 months at once. Read alone the per-block columns could not answer the question they appear to answer: a block one third the length of the sample carries roughly 58% of the full-sample statistic (Section 2.6), so three blocks that fail to reject are the expected outcome of a perfectly stable effect. Here, though, they do not *all* fail to reject: the first block rejects decisively, the last does not, and the across-block test agrees with them.

The main specification's contribution is not stable: it declines monotonically across the sample, from 8.33% in 2010–2014, where on its own it is significant well beyond conventional levels (DM = 3.77), to 3.95% in the second block and 1.68% in 2020–2023, where there is nothing left to test ($p = 0.674$). The decline is not only visual. On the scale of the R^2 columns the last-minus-first contrast is $t = -2.17$ ($p = 0.032$), with the equal-means Wald over the three blocks at $p = 0.094$; on the raw loss scale the contrast has the same sign and falls short of significance ($t = -1.2$, $p = 0.231$). The full-sample 4.22% is therefore an average over a declining profile, and should be read as one.

Two of the implied channels behave differently, which is what makes the decline informative rather than merely discouraging. The penalized model's implied block falls from 0.43% to −3.23%, the same direction as the premium's ($t = -1.54$ on the share scale). The implied betas run the other way: 0.48% in the first block against 1.88% in the last ($t = 1.43$ on the share scale, 1.88 on the loss scale), significant in the last block ($p = 0.009$) and not in the first. Whatever is fading is the priced-correlation signal specifically, not option-implied information in general: the same pairing Section 4.1 volunteers, tabulated here in full.

We cannot say from this sample whether time or turbulence is the operative variable, and we do not claim to. The high-variance regime of Section 4.2 is concentrated in the final block (32 of its 56 months fall there, against 18 in the first block), so the two cuts are reading overlapping months from different angles, and both find the premium's proportional contribution smallest late and in turbulence. Separating a maturing options market from a turbulent final block would take a longer sample than 168 months. What the two cuts jointly establish is narrower and still worth having: the headline is not driven by the crisis months a reader would suspect, and it is not uniform across the window either.

6.2. Where the gain sits in the size distribution

Section 2.6 sets out why the size question is put as a profile rather than as a single value-weighted average. Table 12 reports it: the equal-weighted R^2_{OOS} within terciles of pair size $\frac{1}{2}(ME_{i,t} + ME_{j,t})$, cut within each month, with every model scored against its own baseline on the same tercile. Unlike the calendar blocks, this cut costs no power at all: it splits the cross-section, not the calendar, so each tercile still spans all 168 response months and its test has the length of the headline's.

The gain is flat across size, and significant everywhere. The main specification contributes 4.34%, 4.13% and 4.18% from the smallest to the largest tercile, with p-values of 0.012, 0.023 and 0.018. The forecasting result is therefore not an artifact of small, illiquid names: it is present, and established, among pairs of the largest firms in the index, the pairs an investor could actually trade at scale and the ones a value-weighted statistic would have been dominated by. The large-minus-small contrast is $t = -2.08$ on the share scale ($p = 0.039$) and $t = -0.09$ on the loss scale: proportionally

Table 11
Out-of-sample forecast evaluation within calendar blocks. Each model's marginal contribution and its modified Diebold–Mariano p-value within three calendar blocks of the 168 response months, cut on calendar boundaries (60, 60 and 48 months). Columns and the own-baseline convention are those of Table 3; the final two are the last-minus-first contrast of Equation (20), estimated on all 168 months at once, on the share scale of Equation (21) and on the raw loss scale. Blocks are assigned by response month, and the per-block tests use 3 Newey–West lags against 4 for the full-sample contrasts.

	2010–2014		2015–2019		2020–2023		Last−first t	
Model	ΔR^2 (%)	DM p	ΔR^2 (%)	DM p	ΔR^2 (%)	DM p	share	loss
Panel A: HAR								
HAR+CRP	8.33	0.000	3.95	0.239	1.68	0.674	-2.17	-1.20
HAR+CRP-MW	0.43	0.584	2.54	0.219	-1.96	0.674	-1.16	-0.47
HAR+VRP	0.20	0.750	1.54	0.032	-10.36	0.156	-1.65	-1.48
HAR+skew	-0.22	0.660	-0.18	0.174	-0.11	0.190	0.65	0.19
HAR+$\beta^{\mathbb{Q}}$	0.48	0.177	2.44	0.000	1.88	0.009	1.43	1.88
Panel B: SHAR								
SHAR	0.52	0.332	0.25	0.648	0.92	0.253	-0.12	0.55
SHAR+CRP	7.76	0.001	3.57	0.243	1.11	0.761	-2.15	-1.29
SHAR+CRP-MW	-0.04	0.869	2.15	0.238	-2.52	0.572	-1.14	-0.54
SHAR+VRP	0.32	0.580	1.19	0.066	-8.70	0.139	-1.73	-1.57
SHAR+skew	-0.22	0.687	-0.20	0.225	-0.08	0.428	0.71	0.24
SHAR+$\beta^{\mathbb{Q}}$	0.48	0.169	2.30	0.000	1.74	0.006	1.35	1.86
Panel C: SHAR-F								
SHAR-F	1.85	0.035	-0.07	0.921	2.08	0.077	-0.28	0.26
SHAR-F+CRP	6.93	0.001	3.64	0.233	0.39	0.910	-1.96	-1.22
SHAR-F+CRP-MW	-0.05	0.852	2.63	0.165	-2.75	0.565	-1.11	-0.54
SHAR-F+VRP	0.50	0.294	1.18	0.070	-9.74	0.139	-1.71	-1.59
SHAR-F+skew	-0.44	0.457	-0.19	0.292	-0.07	0.480	0.92	0.63
SHAR-F+$\beta^{\mathbb{Q}}$	0.22	0.447	1.90	0.000	1.32	0.028	1.50	1.61
Panel D: SHAR-F-L								
SHAR-F-L	11.89	0.002	7.18	0.136	6.11	0.241	-0.82	-0.76
SHAR-F-L+CRP	1.05	0.478	-0.86	0.542	-1.82	0.544	-1.57	-0.87
SHAR-F-L+CRP-MW	-1.72	0.257	-0.60	0.317	-4.42	0.123	-1.40	-1.00
SHAR-F-L+VRP	0.92	0.145	1.31	0.066	-5.74	0.108	-2.49	-1.87
SHAR-F-L+skew	-0.60	0.120	0.17	0.316	-0.75	0.194	0.07	-0.41
SHAR-F-L+$\beta^{\mathbb{Q}}$	-0.09	0.790	0.79	0.003	0.69	0.339	1.08	0.98
Panel E: LASSO								
LASSO (realized)	10.54	0.000	9.71	0.022	7.78	0.134	-0.10	-0.29
LASSO (realized+implied)	0.43	0.737	2.37	0.015	-3.23	0.107	-1.54	-1.42

a shade larger in small pairs, negligibly so in levels, against a level that clears significance in all three terciles. The implied betas tilt the other way ($t = 1.97$, $p = 0.05$), the only row in the grid whose contribution is reliably larger among large pairs.

This is also the cut that decides how the forecasting result and the pairs-trading result of Section 5.3 fit together. They are consistent: the spread's value-weighted table tells the same story as its equal-weighted one (Tables 8 and 9), and the forecast that conditions it is equally accurate across the size distribution. Neither result is a small-cap phenomenon.

7. Conclusion

Option-implied signals do improve out-of-sample forecasts of monthly realized correlation, and this paper's central result is what that improvement turns out to be made of. Augmenting the HAR benchmark with the pairwise and market correlation risk premia raises the out-of-sample R^2 by 4.22% ($p = 0.015$) over 168 response months and 14,144,670 pair-months: a genuine gain, on the strictest common-sample design we could build. Two identification

Table 12
Out-of-sample forecast evaluation within terciles of pair size. Each model's marginal contribution and its modified Diebold–Mariano p-value within terciles of average pair capitalization $\frac{1}{2}(ME_{i,t} + ME_{j,t})$, cut within each month so the terciles are cross-sectional (Section 2.6); 4714940, 4714865 and 4714865 pair-months. Columns and the own-baseline convention are those of Table 3. Every tercile spans all 168 response months, so unlike the calendar blocks of Table 11 the per-tercile tests pay no power cost. The final two columns are the large-minus-small contrast, taken as the paired difference of the two terciles' monthly loss-differential series (the cross-sectional counterpart of Equation (20), which a partition of months cannot pair), on the share scale of Equation (21) and on the raw loss scale.

	Small		Mid		Large		Large−small t	
Model	ΔR^2 (%)	DM p	ΔR^2 (%)	DM p	ΔR^2 (%)	DM p	share	loss
Panel A: HAR								
HAR+CRP	4.34	0.012	4.13	0.023	4.18	0.018	-2.08	-0.09
HAR+CRP-MW	0.16	0.933	0.60	0.712	0.70	0.679	-0.84	0.71
HAR+VRP	-3.85	0.269	-2.56	0.302	-2.24	0.362	0.43	1.25
HAR+skew	-0.24	0.064	-0.20	0.208	-0.06	0.837	-0.58	1.07
HAR+$\beta^{\mathbb{Q}}$	1.60	0.004	2.03	0.000	1.72	0.000	1.97	0.70
Panel B: SHAR								
SHAR	0.65	0.124	0.58	0.128	0.41	0.236	-1.50	-0.92
SHAR+CRP	3.88	0.016	3.61	0.031	3.69	0.023	-2.04	-0.14
SHAR+CRP-MW	-0.35	0.822	0.13	0.929	0.29	0.858	-0.67	0.78
SHAR+VRP	-3.27	0.258	-2.16	0.282	-1.81	0.361	0.51	1.23
SHAR+skew	-0.24	0.081	-0.19	0.263	-0.06	0.839	-0.58	1.05
SHAR+$\beta^{\mathbb{Q}}$	1.48	0.005	1.90	0.000	1.64	0.000	2.05	0.77
Panel C: SHAR-F								
SHAR-F	1.36	0.023	1.09	0.029	0.95	0.031	-0.62	-1.01
SHAR-F+CRP	3.61	0.028	3.19	0.060	3.16	0.056	-2.45	-0.46
SHAR-F+CRP-MW	-0.17	0.917	0.31	0.850	0.35	0.842	-0.92	0.57
SHAR-F+VRP	-3.67	0.257	-2.41	0.282	-2.05	0.357	0.63	1.25
SHAR-F+skew	-0.29	0.057	-0.22	0.211	-0.11	0.613	-0.45	1.06
SHAR-F+$\beta^{\mathbb{Q}}$	1.05	0.026	1.53	0.000	1.34	0.000	2.22	1.33
Panel D: SHAR-F-L								
SHAR-F-L	9.63	0.000	7.93	0.002	6.32	0.015	-3.72	-2.82
SHAR-F-L+CRP	-0.57	0.862	-0.77	0.683	-0.92	0.577	-2.18	-1.02
SHAR-F-L+CRP-MW	-2.66	0.032	-2.02	0.073	-1.94	0.080	-0.53	0.79
SHAR-F-L+VRP	-2.08	0.348	-0.90	0.530	-0.81	0.618	0.79	1.25
SHAR-F-L+skew	-0.39	0.065	-0.37	0.102	-0.24	0.245	-0.10	1.06
SHAR-F-L+$\beta^{\mathbb{Q}}$	0.14	0.783	0.76	0.017	0.75	0.009	0.54	1.64
Panel E: LASSO								
LASSO (realized)	10.25	0.000	9.24	0.000	8.29	0.001	-2.62	-1.52
LASSO (realized+implied)	-0.07	0.830	0.16	0.723	-0.20	0.963	-1.48	-0.25

arms then attribute it. Rebuild the premium from Martin–Wagner betas, functions of date-t option prices alone, and the gain is 0.49% ($p = 0.767$); hand the realized benchmark the same year of correlation memory the Buss–Vilkov betas are anchored on, as three trailing means of the pair's own realized correlation, and the benchmark alone earns 7.94% ($p = 0.002$) while the premium's contribution on top of it is −0.76% ($p = 0.701$), with the penalized pair agreeing (−0.04% for the whole implied block, $p = 0.829$). Identified by construction and by control, the premium's forecasting content over HAR is long realized-correlation memory that the HAR family happened to lack. The absorption is selective rather than mechanical: the implied betas keep the grid's one remaining positive margin at that strongest benchmark (0.56%, $p = 0.064$), and a placebo built from sixty-month *physical* betas, features that contribute significantly at the weaker rungs, adds 0.01% there ($p = 0.916$), so the residual, unestablished as it is, is at least not more of the same memory. The attribution travels beyond this design. Any implied-correlation measure that inherits a historical correlation input, as the widely used Buss–Vilkov construction does by design and as Hollstein and

Prokopczuk (2016) document is the reason it performs well, carries that memory into every forecasting comparison whose realized side lacks it, and a benchmark of three trailing means is enough to find out how much of the measured "option-implied" gain is that inheritance.

What the option market does contribute, no realized substitute in our design reproduces, and it took a placebo to establish it. A covariance matrix assembled from $N(N-1)/2$ pooled pairwise forecasts is frequently not positive definite, the failure prices (a Spearman rank correlation of 0.88 between failure frequency and realized portfolio risk across the grid), and adding the correlation risk premium *repairs* it: the only significant reductions in realized minimum-variance portfolio risk in the unconstrained grid (Diebold–Mariano statistics of 3.76 and 2.71), significant on every premium cell once the market factor is constrained out of both sides (2.33, 4.3, 3.78). The placebo is what makes this an option-market finding rather than matrix algebra: pair features built from sixty-month *physical* betas match the premium's forecasting content (+2.05% over SHAR, $t = 4.87$) and actively destroy coherence (raw failure rates of 26.2% against the premium's 11.9% at the same rung; a realized-variance differential of −2.66 where the premium's is +3.76), while the history-free premium breaks the matrix outright. The option content is necessary for the repair, not sufficient; the repairing object is the premium as the option market's correlation view is actually assembled. And the repair earns no fee: with the ten-basis-point charge applied in the premium's favor, the switch still costs a mean-variance investor −0.14% to −0.78% per year, the market-neutral variant's small positive fee (0.35% to 0.19%) being the one qualification that attaches to it.

The remaining margins say the same thing, each with its mechanism attached and every return charged. Collapsed into an average predicted correlation, the forecasts predict the equity premium no better and no sharper than their realized baselines ($t = 1.92$ against 2.01), the purely forward-looking market premium predicts nothing ($t = 0.31$), and the macroeconomic controls absorb what the average correlation carried: the cross-sectional content of a pairwise forecast is averaged away by construction. Read as a cross-sectional ranking, the one margin that neither aggregates nor inverts, the point estimates order as forecast accuracy predicts and no cell reaches significance (2.91%, $t = 1.27$, for the main specification; largest $|t|$ 1.92): the smallest alpha the window could have established is 4.57%, above the HAR-conditioned spread of Bollerslev et al. (2026), and the break-even cost of 6.8 basis points prices a gross alpha indistinguishable from zero. Asked only to measure the risk of characteristic-sorted long-short portfolios, every model in the grid understates it (mean ratios of 0.63 to 0.85 against a target of one) and the premium-augmented models understate most, because a pooled forecast compresses the cross-sectional dispersion of correlations, and the premium's common structure compresses it further. The two portfolio mechanisms are one fact with two signs: common structure is what a covariance matrix needs and what a dispersion bet cannot afford. Our nulls also sit consistently in the published record: DeMiguel et al. (2013) find that option-implied *correlations* do not improve portfolio selection even as their implied volatilities and skewness do, attributing that failure to an unstable matrix, and Kempf et al. (2015) report minimum-variance portfolios built on implied correlations alone underperforming. Both test the option market's correlation view as a level swapped into a historical matrix, and that instability is the incoherence Section 5.1 measures and prices. The record and our charged grid agree at the boundary; what the decomposition adds is attribution: the failure belongs to the level's form, not to the option content that, anchored on realized structure, repairs the matrix instead.

Four limits bound what we have shown. First, the evidence is one market and one horizon: S&P 500 membership over 2005–2023, at a monthly horizon matched to 30-day options, with the sample bounded at 2023, where our option-implied data end. Second, the evaluation cannot begin until the rolling window is consumed, so it excludes the 2008 crisis: the episode where out-of-sample equity-premium predictability concentrates (Rapach et al., 2010), and the reason our aggregate null is a post-crisis statement, read against Buss et al. (2019), whose crisis-inclusive, longer-horizon design does find prediction. Third, our implied correlation is a single-index construction reading only the second moment of the risk-neutral dependence structure (Bondarenko and Bernard, 2024), and the coherence result is identified only as far as its two control arms reach, against physical rank-one structure and against the history-free premium, not against every conceivable repair. Fourth, the calendar decline of the premium's contribution (8.33% to 1.68%) cannot be separated, on 168 months, into time, turbulence, or the memory channel; we report it without choosing.

Each limit names its extension: estimators that impose positive definiteness by construction (Chiriac and Voev, 2011) or model the measurement error in realized covariances (Bollerslev et al., 2018) would test whether the coherence value survives better assembly; a richer risk-neutral dependence measure, a crisis-reaching window, and markets beyond US equities each press a different edge. What the paper establishes is sharper than either a confirmation or a simple null. The correlation risk premium's forecasting gain over HAR is real and is realized memory; the option

market's own contribution to portfolio choice is coherence, not accuracy; and on this sample neither earns a fee. Whether an estimator can impose what the option market reveals, common structure across thousands of forecasts, without the option data is the question the boundary leaves open.

Declaration of competing interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data availability

The option-implied datasets are publicly available from the sources cited in Section 3. The intraday price data were obtained under a commercial license from FirstRate Data and cannot be redistributed. Code reproducing all results from the raw inputs is available from the author.

References

An, B.J., Ang, A., Bali, T.G., Cakici, N., 2014. The joint cross section of stocks and options. Journal of Finance 69, 2279–2337.

Ang, A., Chen, J., 2002. Asymmetric correlations of equity portfolios. Journal of Financial Economics 63, 443–494. doi:10.1016/S0304-405X(02)00068-5.

Audrino, F., Corsi, F., 2010. Modeling tick-by-tick realized correlations. Computational Statistics & Data Analysis 54, 2372–2382. doi:10.1016/j.csda.2009.09.033.

Bakshi, G., Kapadia, N., 2003a. Delta-hedged gains and the negative market volatility risk premium. Review of Financial Studies 16, 527–566.

Bakshi, G., Kapadia, N., 2003b. Volatility risk premiums embedded in individual equity options: Some new insights. Journal of Derivatives 11, 45–54.

Bakshi, G., Kapadia, N., Madan, D., 2003. Stock return characteristics, skew laws, and the differential pricing of individual equity options. Review of Financial Studies 16, 101–143.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. Econometrica 72, 885–925. doi:10.1111/j.1468-0262.2004.00515.x.

Bollerslev, T., Li, J., Patton, A.J., Quaedvlieg, R., 2020. Realized semicovariances. Econometrica 88, 1515–1551.

Bollerslev, T., Li, S.Z., Tang, Y., 2026. Forecasting and managing correlation risks. Management Science doi:10.1287/mnsc.2024.08294. articles in Advance, published online 24 April 2026.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2018. Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. Journal of Econometrics 207, 71–91. doi:10.1016/j.jeconom.2018.05.004.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2022. Realized semibetas: Disentangling "good" and "bad" downside risks. Journal of Financial Economics 144, 227–246.

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Review of Financial Studies 22, 4463–4492. doi:10.1093/rfs/hhp008.

Bollerslev, T., Todorov, V., 2011. Tails, fears, and risk premia. Journal of Finance 66, 2165–2211.

Bondarenko, O., Bernard, C., 2024. Option-implied dependence and correlation risk premium. Journal of Financial and Quantitative Analysis 59, 3139–3189. doi:10.1017/S0022109023000960.

Britten-Jones, M., Neuberger, A., 2000. Option prices, implied price processes, and stochastic volatility. Journal of Finance 55, 839–866.

Buraschi, A., Kosowski, R., Trojani, F., 2014. When there is no place to hide: Correlation risk and the cross-section of hedge fund returns. Review of Financial Studies 27, 581–616. doi:10.1093/rfs/hht070.

Buss, A., Schönleber, L., Vilkov, G., 2019. Expected correlation and future market returns. Working paper, SSRN 3114063; CEPR Discussion Paper 12760.

Buss, A., Vilkov, G., 2012. Measuring equity risk with option-implied correlations. Review of Financial Studies 25, 3113–3140.

Buss, A., Vilkov, G., 2020. Data for "measuring equity risk with option-implied correlations". Data set.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Carr, P., Wu, L., 2009. Variance risk premiums. Review of Financial Studies 22, 1311–1341. doi:10.1093/rfs/hhn038.

Chang, B.Y., Christoffersen, P., Jacobs, K., Vainberg, G., 2012. Option-implied measures of equity risk. Review of Finance 16, 385–428.

Chen, A.Y., Zimmermann, T., 2022. Open source cross-sectional asset pricing. Critical Finance Review 11, 207–264.

Chen, H., Chen, S., Chen, Z., Li, F., 2019. Empirical investigation of an equity pairs trading strategy. Management Science 65, 370–389.

Chiriac, R., Voev, V., 2011. Modelling and forecasting multivariate realized volatility. Journal of Applied Econometrics 26, 922–947. doi:10.1002/jae.1152.

Christoffersen, P., Errunza, V., Jacobs, K., Langlois, H., 2012. Is the potential for international diversification disappearing? a dynamic copula approach. Review of Financial Studies 25, 3711–3751. doi:10.1093/rfs/hhs104.

Christoffersen, P., Jacobs, K., Chang, B.Y., 2013. Forecasting with option-implied information, in: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Elsevier. volume 2A, pp. 581–656.

Christoffersen, P.F., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862. doi:10.2307/2527341.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

DeMiguel, V., Garlappi, L., Uppal, R., 2009. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? Review of Financial Studies 22, 1915–1953. doi:10.1093/rfs/hhm075.

DeMiguel, V., Plyakha, Y., Uppal, R., Vilkov, G., 2013. Improving portfolio selection using option-implied volatility and skewness. Journal of Financial and Quantitative Analysis 48, 1813–1845. doi:10.1017/S0022109013000616.

Dennis, P., Mayhew, S., 2002. Risk-neutral skewness: Evidence from stock options. Journal of Financial and Quantitative Analysis 37, 471–493.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13, 253–263.

Dong, X., Li, Y., Rapach, D.E., Zhou, G., 2022. Anomalies and the expected market return. Journal of Finance 77, 639–681.

Driessen, J., Maenhout, P.J., Vilkov, G., 2009. The price of correlation risk: Evidence from equity options. Journal of Finance 64, 1377–1406.

Driessen, J., Maenhout, P.J., Vilkov, G., 2020. Data for "option-implied correlations and the price of correlation risk". Data set. DOI 10.17605/OSF.IO/CKGYF.

Driessen, J., Maenhout, P.J., Vilkov, G., 2021. Option-implied correlations and the price of correlation risk. Working paper, SSRN 2166829.

Driscoll, J.C., Kraay, A.C., 1998. Consistent covariance matrix estimation with spatially dependent panel data. Review of Economics and Statistics 80, 549–560.

Duan, J.C., Wei, J., 2009. Systematic risk and the price structure of individual equity options. Review of Financial Studies 22, 1981–2006.

Engle, R.F., Colacito, R., 2006. Testing and valuing dynamic correlations for asset allocation. Journal of Business & Economic Statistics 24, 238–253. doi:10.1198/073500106000000017.

Engle, R.F., Figlewski, S., 2015. Modeling the dynamics of correlations among implied volatilities. Review of Finance 19, 991–1018.

Engle, R.F., Kelly, B., 2012. Dynamic equicorrelation. Journal of Business & Economic Statistics 30, 212–228. doi:10.1080/07350015.2011.652048.

Epps, T.W., 1979. Comovements in stock prices in the very short run. Journal of the American Statistical Association 74, 291–298.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E.F., French, K.R., 2020. Comparing cross-section and time-series factor models. Review of Financial Studies 33, 1891–1926.

Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using "realized" volatility. Journal of Financial Economics 67, 473–509.

Gatev, E., Goetzmann, W.N., Rouwenhorst, K.G., 2006. Pairs trading: Performance of a relative-value arbitrage rule. Review of Financial Studies 19, 797–827. doi:10.1093/rfs/hhj020.

Goyal, A., Saretto, A., 2009. Cross-section of option returns and volatility. Journal of Financial Economics 94, 310–326.

Green, J., Hand, J.R.M., Zhang, X.F., 2013. The supraview of return predictive signals. Review of Accounting Studies 18, 692–730.

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

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics 23, 365–380.

Hansen, P.R., Lunde, A., 2005. A realized variance for the whole day based on intermittent high-frequency data. Journal of Financial Econometrics 3, 525–554.

Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting 13, 281–291.

Herskovic, B., Kelly, B., Lustig, H., Van Nieuwerburgh, S., 2016. The common factor in idiosyncratic volatility: Quantitative asset pricing implications. Journal of Financial Economics 119, 249–283.

Hollstein, F., Prokopczuk, M., 2016. Estimating beta. Journal of Financial and Quantitative Analysis 51, 1437–1466.

Jiang, G.J., Tian, Y.S., 2005. The model-free implied volatility and its information content. Review of Financial Studies 18, 1305–1342. doi:10.1093/rfs/hhi027.

Kelly, B., Lustig, H., Van Nieuwerburgh, S., 2016. Too-systemic-to-fail: What option markets imply about sector-wide government guarantees. American Economic Review 106, 1278–1319. doi:10.1257/aer.20120389.

Kempf, A., Korn, O., Saßning, S., 2015. Portfolio optimization using forward-looking information. Review of Finance 19, 467–490.

Krishnan, C.N.V., Petkova, R., Ritchken, P., 2009. Correlation risk. Journal of Empirical Finance 16, 353–367. doi:10.1016/j.jempfin.2008.10.005.

Kupiec, P.H., 1995. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 3, 73–84. doi:10.3905/jod.1995.407942.

Ledoit, O., Wolf, M., 2004. Honey, i shrunk the sample covariance matrix. Journal of Portfolio Management 30, 110–119. doi:10.3905/jpm.2004.110.

Longin, F., Solnik, B., 2001. Extreme correlation of international equity markets. Journal of Finance 56, 649–676. doi:10.1111/0022-1082.00340.

Martin, I., 2017. What is the expected return on the market? Quarterly Journal of Economics 132, 367–433.

Martin, I.W.R., Wagner, C., 2019. What is the expected return on a stock? Journal of Finance 74, 1887–1929.

Mueller, P., Stathopoulos, A., Vedolin, A., 2017. International correlation risk. Journal of Financial Economics 126, 270–299. doi:10.1016/j.jfineco.2016.09.012.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Novy-Marx, R., Velikov, M., 2016. A taxonomy of anomalies and their trading costs. Review of Financial Studies 29, 104–147.

Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Review of Economics and Statistics 97, 683–697.

Pollet, J.M., Wilson, M., 2010. Average correlation and stock market returns. Journal of Financial Economics 96, 364–380.

Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862. doi:10.1093/rfs/hhp063.

Rehman, Z., Vilkov, G., 2012. Risk-neutral skewness: Return predictability and its sources. Working paper, SSRN 1301648.

Rehman, Z., Vilkov, G., 2020. Data for "risk-neutral skewness: Return predictability and its sources". Data set. DOI 10.17605/OSF.IO/BTVDH.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Skintzi, V.D., Refenes, A.N., 2005. Implied correlation index: A new measure of diversification. Journal of Futures Markets 25, 171–197. doi:10.1002/fut.20137.

Stambaugh, R.F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104, 288–302.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58, 267–288.

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

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.