Supplementary appendix to "Learn the measure, estimate the moment: machine-learned drivers in dynamic conditional correlation models"

Yongdeng Xu

September 16, 2026

This supplementary appendix reports the theory behind the estimation step of the paper and the results that the paper summarizes or points to. Section S1 states the consistency results of Section 2.5 of the paper, Section S2 proves them and reports the simulation, Section S3 covers the Dow Jones panel and Section S4 the S&P 500 panel. Section, equation and table numbers without an S refer to the paper.

S1 Consistency of the two-step estimator

This section states conditions under which the second step of Section 2.5 of the paper is consistent and the filtered state converges although the driver is estimated. We treat the correlation block, which is the binding case for two reasons. Its criterion is a joint likelihood over the cross-section, so a first-stage error in any pair reaches every element of the state; and the recursion makes the criterion at date t depend on the driver history rather than on a contemporaneous fitted value, which takes the problem outside the standard two-step theory. The volatility block is the easier case (Remark S2).

Throughout, $\phi = (\alpha, \beta)$ and T is the number of periods in the estimation window. $\varepsilon_t = D_t^{-1} r_t$ are the returns standardized by the population variances h_t, and $\hat{\varepsilon}_t = \hat{D}_t^{-1} r_t$ those standardized by the fitted variances $\hat{h}_t$. N_t^0 is the population driver, a measurable function of the features dated $t-1$ and earlier, $\hat{N}_t$ the learned driver and $\nu_t = \|\hat{N}_t - N_t^0\|$, and w_t collects the features from which the drivers and the variances are learned. Norms are Frobenius unless

marked otherwise, $\|\cdot\|_2$ is the spectral norm, and C denotes a finite constant, depending on k and on the constants of the assumptions, that may change from line to line and does not depend on T or on ϕ.

S1.1 The state and the criterion

Every driver of Section 2.3 of the paper is a correlation matrix, so $N_{t-1} - \bar{N}$ and $\mathcal{S}(N_{t-1} - \bar{N})$ have a zero diagonal and the state in (7) of the paper has a unit diagonal whenever R_{t-1} has one. Started at $R_0 = \bar{R}$, it keeps a unit diagonal at every date, and unrolling the recursion gives the state in closed form,

$$R_t(\phi; N) \;=\; \bar{R} + \alpha \sum_{j=0}^{t-1} \beta^j \, \mathcal{S}\big(N_{t-1-j} - \bar{N}\big), \tag{S.1}$$

the target plus a geometrically weighted sum of the driver's deviations from its mean. When $\mathcal{S}$ removes the mean over the pairs, it is the orthogonal projection, in the Frobenius inner product, onto the symmetric matrices with a zero diagonal whose off-diagonal elements have mean zero, so it is linear and $\|\mathcal{S}(A)\| \le \|A\|$; in coordinates, $\mathcal{S}(v) = v - \bar{v}\,\iota$ is the orthogonal projection onto the complement of ι. The proofs use the driver only through the norms of its deviations and of their differences, which $\mathcal{S}$ does not increase, so every bound in the proofs holds for either choice of $\mathcal{S}$. We write

$$Q_T(\phi; N, \varepsilon) \;=\; \frac{1}{2T} \sum_{t=1}^{T} \Big[\log \big| R_t(\phi; N) \big| + \varepsilon_t' R_t(\phi; N)^{-1} \varepsilon_t \Big] \tag{S.2}$$

for the criterion of (9) of the paper, so that $\hat{\phi} = \operatorname{argmin}_{\phi \in \Phi} Q_T(\phi; \hat{N}, \hat{\varepsilon})$ with the state built from the estimated target $\widehat{\bar{R}}$ and driver mean $\widehat{\bar{N}}$, and $Q(\phi)$ for the expectation of the summand of (S.2) at the population driver, whose state is the infinite-history version of (S.1).

S1.2 Assumptions and results

Assumption S1 (Regularity)**.** *(i)* $\{\varepsilon_t, w_t\}$ *is strictly stationary and absolutely regular with geometrically decaying coefficients, with* $\mathrm{E}\|w_t\|^{2+\epsilon} < \infty$ *and* $\mathrm{E}\|\varepsilon_t\|^{2+\epsilon} < \infty$ *for some* $0 < \epsilon \le 2$*. (ii)* Φ *is compact and contained in* $\{\alpha \ge 0,\ \beta \ge 0,\ \alpha + \beta \le 1 - \eta\}$ *for some* $\eta > 0$*, and* ϕ_0 *is interior to* Φ*. (iii) The state is well conditioned:* $\inf_t \lambda_{\min}\{R_t(\phi; N)\} \ge \varsigma > 0$ *uniformly on* Φ*,*

for every driver path that takes values in the set of correlation matrices. (iv) $Q(\phi)$ is uniquely minimized on Φ at ϕ_0.

Assumption S2 (First stage)**.** *The learned drivers converge in mean square, $T^{-1}\sum_t \|\hat{N}_t - N_t^0\|^2 \xrightarrow{p} 0$; the fitted variances satisfy $T^{-1}\sum_t \|\hat{h}_t - h_t\|^2 \xrightarrow{p} 0$, and the population and the fitted variances lie in $[\underline{h}, \overline{h}]$ for some $0 < \underline{h} \le \overline{h} < \infty$, the fitted ones with probability approaching one; and the estimated target and driver mean converge, $\widehat{\bar{R}} \xrightarrow{p} \bar{R}$ and $\widehat{\bar{N}} \xrightarrow{p} \bar{N}$.*

Condition (iii) is a condition on the state and, as Section 2.3 of the paper notes, it is not guaranteed in correlation space. It is the condition the argument uses, and it is met in the application: over the 183 evaluation months of the monthly panel, the smallest eigenvalue of the correlation-space links lies between 0.061 and 0.068 on the five-minute grid and between 0.067 and 0.071 on the fifteen-minute grid (Section S4). Assumption S2 is stated on the empirical average because that is what the criterion sees, and it involves three estimated nuisances besides the driver: the fitted variances, which build the standardized returns, the target and the driver mean.

Theorem S1. *Under Assumptions S1 and S2, $\hat{\phi} \xrightarrow{p} \phi_0$, and the filtered state is consistent in mean square,*

$$\frac{1}{T}\sum_{t=1}^{T} \left\| R_t(\hat{\phi}; \hat{N}) - R_t(\phi_0; N^0) \right\|^2 \xrightarrow{p} 0, \tag{S.3}$$

where $R_t(\hat{\phi}; \hat{N})$ is built from $\widehat{\bar{R}}$ and $\widehat{\bar{N}}$, and $R_t(\phi_0; N^0)$ from $\bar{R}$ and $\bar{N}$.

The coordinate recursion (8) of the paper is covered once the drivers are bounded. Its state is $\hat{\gamma}_t(\phi; m) = \bar{\gamma} + \alpha \sum_{j<t} \beta^j \, \mathcal{S}(m_{t-1-j} - \bar{m})$, with $R_t(\phi; m) = \mathcal{A}^{-1}\{\hat{\gamma}_t(\phi; m)\}$, and Assumptions S1 and S2 are read with the coordinate driver m in place of N, the coordinate target $\bar{\gamma} = \mathrm{vecl}(\log \bar{R})$ in place of $\bar{R}$ and the driver mean $\bar{m}$ in place of $\bar{N}$.

Corollary S1. *Suppose that Assumption S1(i), (ii) and (iv) and Assumption S2 hold for the coordinate recursion, and that the population and the learned coordinate drivers take values in a compact set $\mathcal{K} \subset \mathbb{R}^{k(k-1)/2}$, the learned ones with probability approaching one. Then $\hat{\phi} \xrightarrow{p} \phi_0$ and $T^{-1}\sum_t \|R_t(\hat{\phi}; \hat{m}) - R_t(\phi_0; m^0)\|^2 \xrightarrow{p} 0$.*

In coordinates, condition (iii) is not assumed but implied by the bound on the drivers, because the forecast is positive definite at every argument and the arguments stay in a compact

set. The compact range holds for the drivers we use: the learned coordinates are the outputs of tree ensembles, which are bounded, and the realized coordinates are bounded because the eigenvalues of the realized correlation are floored at 0.02 (Section 3.1 of the paper).

Corollary S2. *Suppose that the conditions of Theorem S1 hold and that the learned driver converges at date T and at each of the last dates of the window, $\nu_{T-j} \xrightarrow{p} 0$ for every fixed $j \geq 0$. Then the one-step forecast converges in probability to the forecast of the link at ϕ_0, the minimizer of Q,*

$$\left\| R_{T+1}(\hat{\phi}; \hat{N}) - R^{\infty}_{T+1}(\phi_0; N^0) \right\| \xrightarrow{p} 0,$$

where $R_{T+1}(\hat{\phi}; \hat{N})$ is the state (S.1) *at date $T+1$ built from $\widehat{\bar{R}}$ and $\widehat{\bar{N}}$, and $R^{\infty}_{T+1}(\phi_0; N^0)$ is the infinite-history state at the population parameters, driver, target and driver mean. Under the conditions of Corollary S1, with $\hat{m}_T \in \mathcal{K}$ with probability approaching one, and $\|\hat{m}_{T-j} - m^0_{T-j}\| \xrightarrow{p} 0$ for every fixed $j \geq 0$, the same holds for the coordinate forecast $\mathcal{A}^{-1}\{\hat{\gamma}_{T+1}(\hat{\phi}; \hat{m})\}$.*

The condition is not implied by Assumption S2, which does not involve the driver at date T and controls the first-stage error only on average over the window, so it allows the error to stay large at a fixed number of dates, while the forecast weights those dates by $\alpha\beta^j$, which does not vanish as T grows. By Markov's inequality the condition holds if $\mathrm{E}\nu^2_{T-j} \to 0$ for every fixed j, as for a prediction formed out of sample by a learner whose L_2 risk converges, which is the rate Remark S4 assumes. The limit is the forecast of the link at its population parameters; it equals the conditional correlation of returns only if the link is correctly specified, which we do not assume.

We give the proofs of Theorem S1 and Corollaries S1 and S2 in Section S2.1.

S1.3 Scope

Remark S1 (The monthly criterion)**.** *On the monthly panel the state is fixed within a month and the criterion is (10) of the paper, whose summand for month m is $n_m \log|R_m| + \mathrm{tr}(R_m^{-1} S_m)$. The argument holds with t indexing months, $\varepsilon_t \varepsilon_t'$ replaced by S_m, and the weight n_m, at most 23 trading days, absorbed into C, with the moment conditions of Assumption S1(i) applied to the daily standardized returns.*

Remark S2 (The volatility block)**.** *The same argument applies to the volatility recursion (5)*

of the paper stock by stock, with the driver $\hat{d}_{it}$ in place of N_{t-1}, the weight δ_i in place of α and the persistence β_i in place of β, on a compact parameter set that bounds β_i away from one and ω_i away from zero, which replaces Assumption S1(iii); the cross-sectional channel is absent, and on the monthly panel the targeted intercept (6) of the paper plays the role of $\bar{R}$. Under the corresponding condition on $\hat{d}_{it}$, it delivers the mean-square convergence of the fitted variances that Assumption S2 takes as given for the correlation block.

Remark S3 (No inference on the link parameters)**.** *Theorem S1 is a consistency statement, and nothing follows from it about the distribution of $\hat{\phi}$. The plug-in score is not orthogonal to the first stage. Writing $m_t(\phi, N)$ for the score of the summand of* (S.2) *and holding the driver mean fixed, the derivative of $\mathrm{E}[m_t(\phi_0, N^0 + \tau\delta)]$ at $\tau = 0$ in the direction of a perturbation δ of the driver path is*

$$\mathrm{E}\Big[\frac{\partial^2 \ell_t}{\partial\phi\,\partial R}\Big(\alpha \sum_{j\geq 0} \beta^j \delta_{t-1-j}\Big)\Big],$$

where ℓ_t is the summand, and it is not zero at the truth: the channel through the score weight is a martingale difference and vanishes, but the curvature channel, which is the information itself, survives. The first-stage error therefore enters the expansion of $\hat{\phi}$ at first order, a standard error computed from the second step alone is not consistent for the variance of $\hat{\phi}$, and no confidence interval for $\hat{\alpha}$ or $\hat{\beta}$ follows from it. Because the pairing is with the driver path rather than with a contemporaneous residual, a correction for generated regressors in a static model (Pagan, 1984) would not repair it.

Remark S4 (What the empirical results rest on)**.** *Removing the curvature term calls for an orthogonal score of the kind constructed by Chernozhukov et al. (2018), which requires a first-stage rate of $o_p(T^{-1/4})$. We do not establish such a rate: the available consistency theory for boosting concerns independent observations and linear or additive function classes, and does not cover trees with early stopping under temporal dependence. For the same reason Assumption S2 is imposed on the empirical average and not derived from primitive conditions on the learner. A cross-fitted first stage, of the kind used for the daily links (Section 3.4 of the paper), delivers it from an L_2 rate of the learner, and that rate is what we assume. We therefore attach no standard error to $\hat{\phi}$ and read $\hat{\alpha}$ as a diagnostic of the weight the returns place on a driver. None of the forecasting claims of Sections 4.1 and 4.2 of the paper rests on the theorem: they compare forecasting methods out of sample in rolling windows, which under Giacomini and White*

(2006) evaluates each method inclusive of its estimation step and requires neither consistency nor asymptotic normality of the parameters inside it. Two limits of these results are explicit. Theorem S1 averages over the dates of one estimation window and Corollary S2 concerns the forecast at the end of one window, whereas the losses average over forecast origins with different windows, which would need a bound uniform over origins; and both concern the one-step forecast, while the multi-step forecasts of Section 4.1 of the paper iterate the link.

S2 Proofs and simulation

S2.1 Proofs

This subsection proves Theorem S1 and Corollaries S1 and S2, in the notation of Section S1, from the unrolled state (S.1), the criterion (S.2) and Assumptions S1 and S2.

Lemma S1 (Geometric collapse)**.** *For every $\phi \in \Phi$ and any two driver paths N and N' with the same mean $\bar{N}$,*

$$\frac{1}{T}\sum_{t=1}^{T}\big\|R_t(\phi;N)-R_t(\phi;N')\big\|^2 \;\le\; \frac{\alpha^2}{(1-\beta)^2}\,\frac{1}{T}\sum_{s=0}^{T-1}\|N_s-N'_s\|^2 \;\le\; \frac{1}{T}\sum_{s=0}^{T-1}\|N_s-N'_s\|^2.$$

Proof. By (S.1) the target and the driver mean cancel, so with $u_s = \|N_s - N'_s\|$, $\|R_t(\phi;N) - R_t(\phi;N')\| \le \alpha \sum_{j<t} \beta^j u_{t-1-j}$. By the Cauchy–Schwarz inequality, $\big(\sum_{j<t}\beta^j u_{t-1-j}\big)^2 \le (1-\beta)^{-1}\sum_{j<t}\beta^j u_{t-1-j}^2$, and averaging over t and exchanging the order of summation, which is permitted because every term is non-negative,

$$\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{j<t}\beta^j u_{t-1-j}\Big)^2 \;\le\; \frac{1}{1-\beta}\sum_{j\ge 0}\beta^j\,\frac{1}{T}\sum_{s=0}^{T-1}u_s^2 \;=\; \frac{1}{(1-\beta)^2}\,\frac{1}{T}\sum_{s=0}^{T-1}u_s^2.$$

Multiplying by α^2 gives the first bound, and $\alpha \le 1-\eta-\beta < 1-\beta$ on Φ gives the second. □

The lemma is where the filter enters. A first-stage error at any date reaches every later state, and the geometric weights stop it from accumulating, uniformly on Φ because Assumption S1(ii) bounds β away from one.

Lemma S2 (Lipschitz criterion)**.** *Let Assumption S1(iii) hold, let N and N' be driver paths that take values in the set of correlation matrices, and write $\Delta_T(\phi) = T^{-1}\sum_t \|R_t(\phi;N) - R_t(\phi;N')\|^2$ and $M_T = \big(T^{-1}\sum_t \|\varepsilon_t\|^{2+\epsilon}\big)^{2/(2+\epsilon)}$. Then, for every $\phi \in \Phi$,*

$$\big|Q_T(\phi;N,\varepsilon) - Q_T(\phi;N',\varepsilon)\big| \;\le\; C\,(1+M_T)\,\max\big\{\Delta_T(\phi)^{1/2},\,\Delta_T(\phi)^{\epsilon/(2+\epsilon)}\big\}.$$

Moreover, $\|R_t(\phi;N) - R_t(\phi';N)\| \le C\|\phi-\phi'\|$ for every t and all $\phi, \phi' \in \Phi$.

Proof. On the convex set $\mathcal{S}_\varsigma = \{S = S' : \lambda_{\min}(S) \ge \varsigma\}$ the maps $S \mapsto \log|S|$ and $S \mapsto u'S^{-1}u$ are continuously differentiable, with derivatives S^{-1} and $-S^{-1}uu'S^{-1}$, whose norms are at

most $k^{1/2}\varsigma^{-1}$ and $\varsigma^{-2}\|u\|^2$. Both states lie in $\mathcal{S}_\varsigma$ by Assumption S1(iii), so the mean-value theorem along the segment joining them bounds the difference of the date-t summands of (S.2) by $(k^{1/2}\varsigma^{-1} + \varsigma^{-2}\|\varepsilon_t\|^2)\, d_t$, with $d_t = \|R_t(\phi; N) - R_t(\phi; N')\| \le 2k$, since a correlation matrix has Frobenius norm at most k. The average of d_t is at most $\Delta_T^{1/2}$. For the average of $\|\varepsilon_t\|^2 d_t$, Hölder's inequality with exponents $(2+\epsilon)/2$ and $(2+\epsilon)/\epsilon$, together with $d_t^{(2+\epsilon)/\epsilon} \le (2k)^{(2-\epsilon)/\epsilon} d_t^2$ for $\epsilon \le 2$, gives the bound $C\, M_T\, \Delta_T^{\epsilon/(2+\epsilon)}$. This establishes the first claim. For the second, by (S.1), $\partial R_t/\partial\alpha = \sum_{j<t} \beta^j\, \mathcal{S}(N_{t-1-j} - \bar{N})$ and $\partial R_t/\partial\beta = \alpha \sum_{j<t} j\beta^{j-1}\, \mathcal{S}(N_{t-1-j} - \bar{N})$, with norms at most $2k(1-\beta)^{-1}$ and $2k\alpha(1-\beta)^{-2}$, hence at most $2k\eta^{-1}$ on the convex set $\{\alpha \ge 0,\ \beta \ge 0,\ \alpha + \beta \le 1 - \eta\}$ that contains Φ, and the mean-value theorem along the segment joining ϕ and ϕ' in that set gives the claim. □

Lemma S3 (Standardized returns)**.** *Under Assumptions S1(i) and S2,* $T^{-1}\sum_t \|\hat{\varepsilon}_t - \varepsilon_t\|^2 \xrightarrow{p} 0$.

Proof. Componentwise, $\hat{\varepsilon}_{it} - \varepsilon_{it} = \varepsilon_{it}\{(h_{it}/\hat{h}_{it})^{1/2} - 1\}$. On the event that the fitted variances lie in $[\underline{h}, \overline{h}]$, whose probability approaches one by Assumption S2, the mean-value theorem gives $|(h_{it}/\hat{h}_{it})^{1/2} - 1| = h_{it}^{1/2}|\hat{h}_{it}^{-1/2} - h_{it}^{-1/2}| \le \frac{1}{2}\,\overline{h}^{1/2}\underline{h}^{-3/2}|\hat{h}_{it} - h_{it}|$, and $\|\hat{h}_t - h_t\| \le k^{1/2}\overline{h}$. Hence, by Hölder's inequality with exponents $(2+\epsilon)/2$ and $(2+\epsilon)/\epsilon$,

$$\frac{1}{T}\sum_t \|\hat{\varepsilon}_t - \varepsilon_t\|^2 \;\le\; C\,\frac{1}{T}\sum_t \|\varepsilon_t\|^2\, \|\hat{h}_t - h_t\|^2 \;\le\; C\, M_T \Big(\frac{1}{T}\sum_t \|\hat{h}_t - h_t\|^2\Big)^{\epsilon/(2+\epsilon)},$$

where the last step bounds the power $2(2+\epsilon)/\epsilon$ of $\|\hat{h}_t - h_t\|$ by a constant times its square. $M_T = O_p(1)$ by Assumption S1(i) and the ergodic theorem, and the last factor is $o_p(1)$ by Assumption S2. □

Proof of Theorem S1. By Theorem 2.1 of Newey and McFadden (1994) it suffices that Φ is compact, that $Q(\phi)$ is continuous and uniquely minimized at ϕ_0, and that the feasible criterion converges to Q uniformly on Φ in probability. Compactness and uniqueness are Assumption S1(ii) and (iv), and continuity follows by dominated convergence, since the summand of (S.2) is continuous in ϕ and, under Assumption S1(iii), dominated by $C(1 + \|\varepsilon_t\|^2)$. For the

uniform convergence we first treat the target and the driver mean as known, and decompose

$$\sup_{\phi\in\Phi}\left|Q_T(\phi;\hat{N},\hat{\varepsilon})-Q(\phi)\right| \;\le\; \underbrace{\sup_{\phi}\left|Q_T(\phi;\hat{N},\hat{\varepsilon})-Q_T(\phi;\hat{N},\varepsilon)\right|}_{(a)}$$
$$+\underbrace{\sup_{\phi}\left|Q_T(\phi;\hat{N},\varepsilon)-Q_T(\phi;N^0,\varepsilon)\right|}_{(b)}+\underbrace{\sup_{\phi}\left|Q_T(\phi;N^0,\varepsilon)-Q(\phi)\right|}_{(c)}.$$

Term (c). The state (S.1) is started at $\bar{R}$, whereas Q uses the infinite-history state. The two differ at date t by $\alpha\sum_{j\ge t}\beta^j\,\mathcal{S}(N^0_{t-1-j}-\bar{N})$, whose norm is at most $2k\beta^t\le 2k(1-\eta)^t$ because $\alpha\le 1-\beta$, so by the per-date bound in the proof of Lemma S2 the two criteria differ by at most $CT^{-1}\sum_t(1-\eta)^t(1+\|\varepsilon_t\|^2)=O_p(T^{-1})$ uniformly on Φ. The criterion with the infinite-history state is the average of a stationary ergodic sequence, since N^0_t is a function of the features dated $t-1$ and earlier, and its summand is continuous in ϕ on the compact Φ and dominated by the integrable $C(1+\|\varepsilon_t\|^2)$; the uniform law of large numbers for such sequences gives $(c)=o_p(1)$.

Term (b). By Lemma S2 and then Lemma S1, uniformly on Φ,

$$(b)\;\le\; C\,(1+M_T)\,\max\left\{V_T^{1/2},\,V_T^{\epsilon/(2+\epsilon)}\right\},\qquad V_T=\frac{1}{T}\sum_s\nu_s^2.$$

$M_T=O_p(1)$ by Assumption S1(i) and $V_T=o_p(1)$ by Assumption S2, so $(b)=o_p(1)$.

Term (a). Only the quadratic form differs, and $\hat{\varepsilon}_t'S^{-1}\hat{\varepsilon}_t-\varepsilon_t'S^{-1}\varepsilon_t=(\hat{\varepsilon}_t-\varepsilon_t)'S^{-1}(\hat{\varepsilon}_t+\varepsilon_t)$ with $\|S^{-1}\|_2\le\varsigma^{-1}$ by Assumption S1(iii), so by the Cauchy–Schwarz inequality

$$(a)\;\le\;\varsigma^{-1}\Big(\frac{1}{T}\sum_t\|\hat{\varepsilon}_t-\varepsilon_t\|^2\Big)^{1/2}\Big(\frac{1}{T}\sum_t\|\hat{\varepsilon}_t+\varepsilon_t\|^2\Big)^{1/2}.$$

The first factor is $o_p(1)$ by Lemma S3, and the second is $O_p(1)$ because $\|\hat{\varepsilon}_t\|\le(\overline{h}/\underline{h})^{1/2}\|\varepsilon_t\|$ with probability approaching one. This channel is not damped by the geometric weights, because it enters the criterion at date t directly rather than through the driver history; it is closed by the bounds on the fitted variances.

The estimated target and driver mean. By (S.1), replacing $(\bar{R},\bar{N})$ by $(\widehat{\bar{R}},\widehat{\bar{N}})$ shifts the state at date t by $(\widehat{\bar{R}}-\bar{R})-\alpha\sum_{j<t}\beta^j\,\mathcal{S}(\widehat{\bar{N}}-\bar{N})$, whose norm is at most $\|\widehat{\bar{R}}-\bar{R}\|+\|\widehat{\bar{N}}-\bar{N}\|$ on

Φ because $\alpha \leq 1-\beta$, and therefore $o_p(1)$ uniformly in t and ϕ by Assumption S2. By Weyl's inequality the shifted states satisfy Assumption S1(iii) with $\varsigma/2$ in place of ς with probability approaching one, so the shift adds a term $o_p(1)$ to (b) through Lemma S2.

The terms give $\sup_\phi |Q_T(\phi; \hat{N}, \hat{\varepsilon}) - Q(\phi)| = o_p(1)$, and hence $\hat{\phi} \xrightarrow{p} \phi_0$. For (S.3), add and subtract $R_t(\phi_0; \hat{N})$ and use $\|x+y\|^2 \leq 2\|x\|^2 + 2\|y\|^2$:

$$\frac{1}{T}\sum_t \left\|R_t(\hat{\phi}; \hat{N}) - R_t(\phi_0; N^0)\right\|^2 \leq \frac{2}{T}\sum_t \left\|R_t(\hat{\phi}; \hat{N}) - R_t(\phi_0; \hat{N})\right\|^2 + \frac{2}{T}\sum_t \left\|R_t(\phi_0; \hat{N}) - R_t(\phi_0; N^0)\right\|^2.$$

The first term is at most $2C^2\|\hat{\phi} - \phi_0\|^2 = o_p(1)$ by the second part of Lemma S2 and the first claim, and the second is $o_p(1)$ by Lemma S1, Assumption S2 and the bound on the shift from the estimated target and driver mean. □

Proof of Corollary S1. The coordinate state is linear in the driver path, so Lemma S1 holds verbatim with $\hat{\gamma}_t$ in place of R_t. On Φ, $\alpha \sum_j \beta^j \leq 1$, so every state satisfies $\|\hat{\gamma}_t - \bar{\gamma}\| \leq 2\sup_{x\in\mathcal{K}} \|x\|$ and lies in a closed ball $\mathcal{K}'$, which is compact and convex; with probability approaching one the same holds, after enlarging the ball, for the states built from $\widehat{\bar{R}}$ and $\widehat{\bar{m}}$, since $\operatorname{vecl}(\log \widehat{\bar{R}}) \xrightarrow{p} \bar{\gamma}$ by the continuity of the matrix logarithm on positive definite matrices. The inverse map is continuously differentiable. Archakov and Hansen (2021) show that for every γ there is a unique vector x such that the symmetric matrix G with off-diagonal elements γ and diagonal x satisfies $\operatorname{diag}\{\exp(G)\} = \iota$, and $\mathcal{A}^{-1}(\gamma) = \exp(G)$. The Jacobian of $x \mapsto \operatorname{diag}\{\exp(G)\}$ is $\int_0^1 \exp(sG) \circ \exp\{(1-s)G\}\, ds$, where $\circ$ is the elementwise product, and it is positive definite by the Schur product theorem, so by the implicit function theorem x, and with it $\mathcal{A}^{-1}$, is continuously differentiable in γ. $\mathcal{A}^{-1}$ is therefore Lipschitz on $\mathcal{K}'$, with a constant L, and $T^{-1}\sum_t \|R_t(\phi; \hat{m}) - R_t(\phi; m^0)\|^2 \leq L^2\, T^{-1}\sum_s \|\hat{m}_s - m_s^0\|^2$. The smallest eigenvalue of $\mathcal{A}^{-1}(\gamma)$ is continuous and positive on $\mathcal{K}'$, so it has a positive minimum there, which gives Assumption S1(iii) for these states; and the state is Lipschitz in ϕ, because $\hat{\gamma}_t$ is, with the bounds of Lemma S2 after replacing $2k$ by $2\sup_{x\in\mathcal{K}} \|x\|$, and $\mathcal{A}^{-1}$ is Lipschitz on $\mathcal{K}'$. The proof of Theorem S1 then applies with these facts in place of Lemmas S1 and S2. □

Proof of Corollary S2. Write $R_t^\infty(\phi; N) = \bar{R} + \alpha \sum_{j\geq 0} \beta^j\, \mathcal{S}(N_{t-1-j} - \bar{N})$ for the infinite-history

state, $R_{T+1}(\phi; N)$ for the state (S.1) at date $T+1$ built from $\bar{R}$ and $\bar{N}$, and $\tilde{R}_{T+1}(\phi; N)$ for the same state built from $\widehat{\bar{R}}$ and $\widehat{\bar{N}}$. By the triangle inequality,

$$\begin{aligned}\big\|\tilde{R}_{T+1}(\hat{\phi}; \hat{N}) - R^{\infty}_{T+1}(\phi_0; N^0)\big\| \;\le\; & \big\|\tilde{R}_{T+1}(\hat{\phi}; \hat{N}) - R_{T+1}(\hat{\phi}; \hat{N})\big\| \\ & + \big\|R_{T+1}(\hat{\phi}; \hat{N}) - R_{T+1}(\phi_0; \hat{N})\big\| \\ & + \big\|R_{T+1}(\phi_0; \hat{N}) - R_{T+1}(\phi_0; N^0)\big\| \\ & + \big\|R_{T+1}(\phi_0; N^0) - R^{\infty}_{T+1}(\phi_0; N^0)\big\|.\end{aligned}$$

The first term is at most $\|\widehat{\bar{R}} - \bar{R}\| + \|\widehat{\bar{N}} - \bar{N}\| = o_p(1)$, as in the proof of Theorem S1. The second is at most $C\|\hat{\phi} - \phi_0\| = o_p(1)$ by the second part of Lemma S2, whose bound does not depend on the date and so holds at $T+1$, since $\hat{N}_T$, like every driver, is a correlation matrix, and by Theorem S1. The fourth is at most $2k\beta_0^{T+1}$, as in term (c). For the third, $\mathcal{S}$ does not increase norms, $\alpha_0 \le 1$, $\beta_0 \le 1-\eta$ and $\nu_s \le 2k$, so for every integer $J \ge 0$

$$\big\|R_{T+1}(\phi_0; \hat{N}) - R_{T+1}(\phi_0; N^0)\big\| \;\le\; \sum_{j=0}^{T} \beta_0^j \, \nu_{T-j} \;\le\; \sum_{j=0}^{J} \nu_{T-j} + \frac{2k(1-\eta)^{J+1}}{\eta}.$$

Given $\delta > 0$, choose J with $2k(1-\eta)^{J+1}/\eta < \delta/2$. The finite sum is $o_p(1)$ by the condition of the corollary, so the third term is below δ with probability approaching one. For the coordinate forecast the four bounds hold for $\hat{\gamma}_{T+1}$, with $2\sup_{x\in\mathcal{K}}\|x\|$ in place of $2k$, which uses $\hat{m}_T \in \mathcal{K}$, and with $\mathrm{vecl}(\log \widehat{\bar{R}}) \xrightarrow{\;p\;} \bar{\gamma}$ in the first term; with probability approaching one every state involved lies in the compact ball $\mathcal{K}'$ of the proof of Corollary S1, on which $\mathcal{A}^{-1}$ is Lipschitz. Hence the coordinate forecast converges as well. □

S2.2 Simulation

A simulation knows the truth, so it can score a selection rule against the true correlation rather than against the realized one. The true correlation of month t comes from a three-factor covariance $\Sigma_t = L_t L_t' + \mathrm{diag}(d)$ with loadings following a stationary VAR(1), $L_t = \bar{L} + 0.98(L_{t-1} - \bar{L}) + 0.05\,\eta_t$, idiosyncratic variances drawn from $U[0.4, 1.2]$ and market loadings with mean 0.75, so that correlations average about 0.30, as they do across S&P 500 stocks. Each month has $D = 21$ days of n intraday efficient returns drawn from $N(0, R_t/n)$, and the observed returns carry two distortions in the *noisy* cells: microstructure noise, an independent

$N(0, 6.4 \times 10^{-4})$ error on every log price, and stale prices, each stock's price being refreshed at a grid point with a probability drawn from $U[0.6, 1]$ and at the close with certainty. The *clean* cells switch both off. We set $k \in \{30, 100, 300\}$ and $n \in \{27, 78\}$, BLT's fifteen-minute grid and a five-minute grid, and run 300 months with five replications per cell: the first 150 months are discarded, the next 60 provide the history for the first selection, and the last 90 are evaluated.

The target is the nonlinearly shrunk correlation of the trailing 48 months of daily returns. Two drivers are compared: the lagged realized correlation, and a pooled projection of it on its lagged value, its means over the last three and twelve months, the two stocks' average correlations with the other stocks and the cross-sectional mean, fitted by least squares on the trailing 120 months, which stands in for the class of Bollerslev et al. (2026). Both enter the recursion (7) of the paper in correlation space, where validity can fail, and (α, β) is re-selected at every evaluation month on the trailing 36 months over a grid of 83 points by one of four rules: BLT's pooled squared error against the realized correlation, the returns likelihood (10) of the paper, and two infeasible oracles that minimize QLIKE and the Frobenius distance against the true correlation.

The returns likelihood selects the weight the truth selects. Table 1 of the paper shows that the QML weight is within 0.04 of the QLIKE oracle's in all 24 combinations of cell and driver, and that both fall with the dimension: for the projection driver in the noisy cells with $n = 27$ the QML weight is 0.63, 0.30 and 0.20 at $k = 30$, 100 and 300. BLT's criterion does not follow this pattern. It selects 0.77 to 0.96 for the projection in the noisy cells, and 0.90 to 0.93 in the clean ones, at every k, which is the prediction of Section 2.6 of the paper: a pairwise criterion is blind to the spectral cost of a weight.

Table S1 reports the levels of the simulated second moments. In the noisy cells the realized correlation averages 0.18 to 0.21 against a true correlation of 0.29 to 0.33, while daily returns keep the true level (0.28 to 0.32), and the realized variance is 1.04 to 1.12 times the variance of daily returns. In the clean cells the realized correlation is within 0.001 of the truth.

Table S2 reports the forecasts in the noisy cells. The forecasts fitted to the measure, the projection and the lagged realized correlation, attain, in every noisy cell, the smallest squared error against the realized correlation and the two largest against the truth: at $k = 30$ and $n = 78$ their true QLIKE is 3.29 and 3.24, against 1.25 for the QML-selected link. At $k = 100$

the links selected on BLT's criterion are indefinite in 2.0% to 40.0% of months across the eight combinations of n, cell and driver; with the projection driver in the noisy cell with $n = 78$ the share is 4.7% and the true QLIKE is 17.10, against no indefinite month and 7.07 under QML. At $k = 300$ the links selected on BLT's criterion, and on the Frobenius oracle, are indefinite in every one of the 90 evaluation months of every cell, with a mean smallest eigenvalue between -0.65 and -0.97, while the QML-selected links are positive definite in every month, with a true QLIKE of 24.2 against 35.4 for the projection at $n = 78$ in the noisy cell.

The clean cells show when fitting the measure pays. When the realized measure is unbiased and k is at most 100, the projection has the lowest true QLIKE in the four clean cells, 0.55 against 0.87 for the QML-selected link at $k = 30$ and $n = 27$. The projection also attains the lowest true risk of the minimum variance portfolio in the four $k = 300$ cells, 13.69 against 14.26 for the QML-selected link in the noisy cell with $n = 78$, so the gain on the likelihood does not translate into every economic criterion.

Table S1 – Simulation: levels

k	n	Cell	True	Realized	Daily returns	RV / daily variance
30	27	clean	0.318	0.318	0.310	1.02
30	27	noisy	0.327	0.207	0.322	1.05
30	78	clean	0.319	0.319	0.312	1.02
30	78	noisy	0.323	0.199	0.320	1.12
100	27	clean	0.310	0.309	0.302	1.02
100	27	noisy	0.312	0.203	0.306	1.05
100	78	clean	0.294	0.293	0.286	1.02
100	78	noisy	0.292	0.178	0.283	1.12
300	27	clean	0.297	0.297	0.291	1.01
300	27	noisy	0.294	0.193	0.287	1.04
300	78	clean	0.303	0.303	0.299	1.01
300	78	noisy	0.300	0.181	0.295	1.11

Post-burn means over five replications. True, realized and daily returns are mean correlations; RV / daily variance is the realized variance divided by the variance of daily returns.

Table S2 – Simulation: forecasts in the noisy cells

Forecast	Selection	MSE, realized	MSE, true	QLIKE, true	Frobenius, true	GMV true risk	Indefinite (%)	Mean correlation
$k = 30$, $n = 27$, *noisy*								
Returns target		0.0279	0.0129	2.66	3.33	43.43	0.0	0.332
Projection		0.0027	0.0238	2.83	4.52	43.16	0.0	0.213
Lagged realized		0.0037	0.0238	3.00	4.52	43.33	0.0	0.214
Link on projection	realized	0.0241	0.0053	1.53	2.11	41.46	0.0	0.336
Link on projection	QML	0.0242	0.0047	1.43	2.01	41.15	0.0	0.337
Link on projection	oracle	0.0242	0.0048	1.43	2.03	41.16	0.0	0.337
Link on lagged realized	realized	0.0239	0.0052	1.60	2.11	41.58	0.0	0.335
Link on lagged realized	QML	0.0240	0.0049	1.51	2.05	41.34	0.0	0.336
Link on lagged realized	oracle	0.0241	0.0052	1.50	2.11	41.39	0.0	0.336
$k = 30$, $n = 78$, *noisy*								
Returns target		0.0264	0.0122	2.53	3.25	40.38	0.0	0.320
Projection		0.0013	0.0254	3.29	4.66	40.75	0.0	0.202
Lagged realized		0.0016	0.0250	3.24	4.62	40.79	0.0	0.202
Link on projection	realized	0.0229	0.0044	1.32	1.94	38.34	0.0	0.324
Link on projection	QML	0.0231	0.0040	1.25	1.87	38.26	0.0	0.325
Link on projection	oracle	0.0231	0.0041	1.25	1.87	38.26	0.0	0.325
Link on lagged realized	realized	0.0229	0.0044	1.37	1.94	38.44	0.0	0.324
Link on lagged realized	QML	0.0230	0.0042	1.26	1.90	38.34	0.0	0.325
Link on lagged realized	oracle	0.0230	0.0042	1.26	1.90	38.34	0.0	0.325
$k = 100$, $n = 27$, *noisy*								
Returns target		0.0235	0.0127	10.32	11.18	24.27	0.0	0.309
Projection		0.0027	0.0205	10.43	14.21	23.61	0.0	0.204
Lagged realized		0.0037	0.0204	15.56	14.19	24.95	0.0	0.205
Link on projection	realized	0.0189	0.0046	24.90	6.72	31.57	4.9	0.309
Link on projection	QML	0.0201	0.0075	7.50	8.61	23.28	0.0	0.309
Link on projection	oracle	0.0201	0.0075	7.50	8.61	23.28	0.0	0.309
Link on lagged realized	realized	0.0188	0.0046	30.40	6.74	36.23	20.2	0.309
Link on lagged realized	QML	0.0201	0.0077	7.63	8.69	23.63	0.0	0.309
Link on lagged realized	oracle	0.0201	0.0077	7.63	8.69	23.63	0.0	0.309
$k = 100$, $n = 78$, *noisy*								
Returns target		0.0236	0.0129	10.47	11.30	23.17	0.0	0.285
Projection		0.0013	0.0224	10.86	14.86	22.74	0.0	0.178
Lagged realized		0.0016	0.0220	11.58	14.73	23.20	0.0	0.179
Link on projection	realized	0.0190	0.0044	17.10	6.55	27.14	4.7	0.285
Link on projection	QML	0.0198	0.0067	7.07	8.13	22.18	0.0	0.285
Link on projection	oracle	0.0198	0.0067	7.07	8.11	22.18	0.0	0.285
Link on lagged realized	realized	0.0190	0.0044	22.59	6.55	27.93	12.4	0.285
Link on lagged realized	QML	0.0198	0.0068	7.22	8.19	22.39	0.0	0.285
Link on lagged realized	oracle	0.0199	0.0069	7.21	8.22	22.39	0.0	0.285
$k = 300$, $n = 27$, *noisy*								
Returns target		0.0214	0.0132	31.27	34.44	14.35	0.0	0.285
Projection		0.0028	0.0190	39.34	41.27	13.60	0.2	0.193
Lagged realized		0.0037	0.0192	189.06	41.39	17.94	0.0	0.193
Link on projection	realized	0.0165	0.0047	–	20.61	–	100.0	0.286
Link on projection	QML	0.0194	0.0104	25.70	30.52	14.00	0.0	0.285
Link on projection	oracle	0.0194	0.0104	25.70	30.51	14.00	0.0	0.285
Link on lagged realized	realized	0.0164	0.0048	–	20.66	–	100.0	0.286
Link on lagged realized	QML	0.0196	0.0108	26.26	31.17	14.28	0.0	0.285
Link on lagged realized	oracle	0.0195	0.0108	26.25	31.08	14.30	0.0	0.285
$k = 300$, $n = 78$, *noisy*								
Returns target		0.0246	0.0127	30.97	33.80	14.51	0.0	0.295
Projection		0.0013	0.0232	35.40	45.60	13.69	0.0	0.181
Lagged realized		0.0016	0.0230	49.48	45.42	14.71	0.0	0.180
Link on projection	realized	0.0203	0.0044	–	19.87	–	100.0	0.295
Link on projection	QML	0.0227	0.0097	24.21	29.53	14.26	0.0	0.295
Link on projection	oracle	0.0227	0.0097	24.21	29.55	14.26	0.0	0.295
Link on lagged realized	realized	0.0202	0.0044	–	19.73	–	100.0	0.295
Link on lagged realized	QML	0.0227	0.0099	24.48	29.78	14.44	0.0	0.295
Link on lagged realized	oracle	0.0227	0.0099	24.48	29.78	14.44	0.0	0.295

MSE, realized and MSE, true are mean squared errors against the realized and the true correlation; QLIKE, Frobenius and the risk of the minimum variance (GMV) portfolio are computed against the true correlation, the risk with unit variances and multiplied by 100. Selection is the rule that selects the weights of a link; indefinite is the share of evaluation months in which the forecast is not positive definite. A dash marks a criterion that is undefined because the forecast is indefinite.

S3 Dow Jones panel

S3.1 All models

Table S3 reports QLIKE against the returns and against the realized covariance for the eleven daily models and three benchmarks at horizons of 1, 5 and 22 days, and Tables S4 and S5 test the comparisons that change one element of the model at a time: the volatility block, the space of the correlation recursion, the driver, and the inverse used at longer horizons. Each model is labelled by its two drivers, as in Section 4.1 of the paper; in correlation space, RV / RC is DCC-GARCH-X. The two criteria give the same sign on every comparison that is significant at the 1% level under either, except the coordinate recursion with linked volatility and the learned driver at 22 days.

S3.2 The volatility block

Replacing the lagged realized variance in the variance equation by the learned driver lowers QLIKE against the returns by 1.16 to 1.23 at one day, 1.33 to 1.35 at five days and 3.25 to 3.51 at 22 days over the four specifications of the correlation block, with every p value below 10^{-4}. Table S6 shows where the gain at longer horizons comes from. The iterated forecast on the lagged realized variance over-predicts the variance of returns more as the horizon lengthens: the mean squared return over the h days is 0.95, 0.87 and 0.68 of its forecast variance at 1, 5 and 22 days. The linked forecast stays close to the level of returns, at 1.06, 0.98 and 0.92, because at longer horizons its recursion is driven by a direct forecast of the h-day average realized variance. A level correction estimated on the trailing 250 origins does not improve it: it raises QLIKE of the linked forecast against the returns by 0.27 at one day ($p = 0.043$), 0.46 at five days ($p < 0.005$) and 1.58 at 22 days ($p = 0.005$).

S3.3 The coordinate space

With the lagged realized correlation as the driver, running the recursion in coordinates lowers QLIKE against the returns by 0.29 to 0.30 at one day, 0.19 to 0.21 at five days and 0.07 to 0.11 at 22 days, with RV and with linked volatility, and every p value is below 0.003. The gain is not a matter of validity, since at $k = 36$ the evaluation's eigenvalue floor binds for no forecast in either space. Nor is it a matter of level, although the coordinate forecasts sit slightly

lower, with a mean implied correlation of 0.304 at one day against 0.307 in correlation space. When the forecasts of both spaces are shifted on trailing data to the level of the realized proxy (Section S3.5), 83% to 94% of the gain remains: 0.24 to 0.25 at one day, 0.17 at five days and 0.08 to 0.09 at 22 days. With the learned driver the coordinate recursion is also better, by between 0.03 and 0.11, and four of the six differences are significant at the 5% level.

S3.4 The learned correlation driver

On the daily panel the learned driver lowers the accuracy of the correlation forecast. Replacing the lagged realized correlation by the learned coordinates raises QLIKE against the returns at one day by 0.55 to 0.58 in correlation space and by 0.75 to 0.82 in coordinates, and at five days by 0.30 to 0.41. At 22 days it raises QLIKE by 0.47 and 0.51 with RV volatility and by 0.24 ($p = 0.006$) and 0.29 ($p = 0.004$) with linked volatility, and every other p value is at most 0.001.

The loss does not come from the generated-regressor problem of Section 2.5. A link estimated on the learner's predictions on its own training rows sees a driver that fits the window better than any prediction fits the days that follow it, which is why our links are estimated on cross-fitted predictions. On these the likelihood puts a mean weight of 0.307 on the learned driver in correlation space and 0.328 in coordinates, against 0.081 and 0.094 on the lagged realized correlation. On a measure whose variances alone are rescaled (Hansen and Lunde, 2005), estimating the links on the in-sample predictions raises these weights from 0.284 and 0.310 to 0.386 and 0.439 and lowers the loss at one day to 0.36 to 0.38 in correlation space and 0.71 in coordinates, while the volatility link gains less, 0.93 to 0.96 at one day, and under-predicts the variance of returns, with a ratio of 1.21. In five windows spread over the sample, cross-fitting lowers the in-window R^2 of the driver against the realized coordinates from between 0.09 and 0.11 to between 0.01 and 0.02. Cross-fitting therefore improves the volatility link and lowers the weight on the learned correlation driver without removing its loss, and on this panel the learned correlation driver adds no information to the lagged realized correlation.

S3.5 The retransformation and the level corrections

Averaging the inverse over draws of the driver's forecast path lowers QLIKE against the returns at 22 days by 0.065 and 0.074 for the lagged realized correlation ($p < 10^{-4}$) and by 0.092 for the

learned driver ($p < 0.001$), less than a fortieth of the volatility link's gain at that horizon. At five days it lowers QLIKE by 0.013 and 0.015 for the lagged realized correlation ($p \leq 0.009$) and does not change it significantly for the learned driver (-0.012, $p = 0.36$). Against the realized covariance the gain is 0.071 and 0.082 at 22 days and 0.015 and 0.017 at five days for the lagged realized correlation, and 0.158 and 0.041 for the learned driver, with every p value below 0.001. The rows marked "simulated" in Table S3 report these forecasts; for the models of Table 3 of the paper the ranking is unchanged. With the lagged realized correlation the simulated forecast lowers the mean implied correlation at 22 days by 0.007.

Table S7 separates this from a level effect. A common shift of the coordinates, estimated on the trailing 250 origins so that the forecasts' mean implied correlation matches the realized proxy's, lowers QLIKE against the returns at one day by 0.08 to 0.16 for the coordinate models and by 0.11 to 0.14 when the same shift is applied to the correlation-space models, where no retransformation is involved, with every p value below 0.02. The shift is about half as large as on a measure whose correlations stay at the intraday level, 0.21 to 0.41, since the rescaling of Section 3.1 already moves the proxy toward the level of returns. At 22 days it lowers QLIKE by 0.01 to 0.13, significantly at the 5% level for three of the six models, two of them in correlation space. The level shift is therefore a calibration of the correlation level that the two spaces share rather than a correction of the retransformation, and the retransformation itself is the small multi-step gain of the simulated inverse.

Against the horizon-matched DCC-GARCH-X, which forecasts each horizon's average directly, the model with linked volatility and the learned coordinate driver is better by 0.87 at one day and by 1.74 and 1.17 at 5 and 22 days ($p < 0.001$). Against the unrestricted DCC-GARCH-X it does not differ significantly at one day (-0.19, $p = 0.37$); the multi-step forecasts of that benchmark are near-singular, with the evaluation's eigenvalue floor binding at 47 of the five-day origins, and are not a usable comparison.

S3.6 Losses against the realized covariance on all origins

Table S8 reports the losses of every Dow Jones model against the realized covariance, rescaled as in Section 3.1 of the paper, on all forecast origins, in QLIKE and in the Frobenius distance, at horizons of 1, 5 and 22 days. On all origins QLIKE ranks the models as in Table S3: the

coordinate recursion with ML-RV and RC attains 47.11, 48.76 and 51.53 at 1, 5 and 22 days, against 48.13, 50.08 and 55.43 for DCC-GARCH-X. The Frobenius distance, a squared error in the units of the covariance that is dominated by the most volatile stocks and days, ranks the models differently at one day. There the recursions on RV and RC (21,660 and 21,771) and the unrestricted DCC-GARCH-X (21,665) are below every model with ML-RV (22,396 to 23,346). At 22 days the models with ML-RV are below DCC-GARCH-X, 10,208 to 10,763 against 11,509, and the horizon-matched DCC-GARCH-X attains the lowest distance, 10,139. The multi-step forecasts of the unrestricted DCC-GARCH-X are near-singular, and their Frobenius distance exceeds 10^6.

S3.7 Economic value

Tables S9 and S10 score the minimum variance portfolios of the models of Table 2 of the paper with the conventions of the S&P 500 panel (Table S17), at horizons of 1, 5 and 22 days. The portfolio is formed from the h-day forecast at the start of each non-overlapping holding period common to every model, 2,352, 470 and 106 periods, and held for h days. Realized risk is the annualized square root of the portfolio variance under the mean rescaled realized covariance of the holding period, returns and the Sharpe ratio are annualized with $252/h$ periods a year and a risk-free rate of 2%, turnover is the sum of the absolute changes in the weights between consecutive holding periods, and the tests and fees are against DCC-GARCH-X. The weights differ from those of the S&P 500 panel in one respect: they are formed from the forecast with the evaluation's eigenvalue floor rather than from the corrected matrix of Bollerslev et al. (2026), and the floor binds for no forecast of the coordinate recursions.

The coordinate ML-RV / RC model attains the lowest realized risk at every horizon, 0.16, 0.25 and 0.65 percentage points below DCC-GARCH-X ($t = -1.84$, -2.54 and -7.60), and the lowest standard deviation of the realized return, 0.16, 0.36 and 1.71 points below ($t = -0.74$, -1.53 and -2.25 for the squared returns). The coordinate recursion on RV and RC alone lowers the realized risk by 0.06, 0.04 and 0.02 points ($t = -5.11$, -2.91 and -1.81), with no significant difference in the squared returns. The horizon-matched DCC-GARCH-X is riskier than DCC-GARCH-X at one and five days, by 0.50 and 0.43 points ($t = 8.20$ and 4.72), and less risky at 22 days, by 0.21 points ($t = -1.70$), where its return standard deviation is 11.26%

and the fee for switching to it is 4.21% a year for $\gamma = 5$ (paired $t = 2.07$), against 3.87% (paired $t = 1.55$) for the ML-RV / RC model. Without the holding period with the largest absolute return difference, the two fees are 3.39% and 4.16%, so the higher fee of the horizon-matched model rests on one holding period. The forecasts of the unrestricted DCC-GARCH-X at $h > 1$ and of its horizon-matched version are near-singular at some origins, and their portfolios are not interpretable. Turnover is lower for the ML-RV / RC model than for DCC-GARCH-X at one and five days, 0.86 and 0.95 against 1.11 and 1.36, and nearly equal at 22 days, 0.98 against 0.99.

Table S3 – Dow Jones panel: QLIKE against returns and against the realized covariance, all models

	Against returns			Against the realized covariance		
Model	$h = 1$	$h = 5$	$h = 22$	$h = 1$	$h = 5$	$h = 22$
Correlation space						
RV / RC (DCC-GARCH-X)	53.47	54.95	58.81	48.13	50.08	55.43
ML-RV / RC	52.31	53.62	55.57	47.34	48.95	51.63
RV / ML-RC	54.05	55.26	59.29	48.66	50.47	56.10
ML-RV / ML-RC	52.86	53.92	55.81	47.83	49.31	52.10
Coordinates						
RV / RC	53.17	54.76	58.75	47.88	49.90	55.35
ML-RV / RC	52.01	53.41	55.46	47.11	48.76	51.53
RV / ML-RC	54.00	55.16	59.26	48.60	50.40	56.17
ML-RV / ML-RC	52.77	53.81	55.75	47.74	49.23	52.16
RV / RC, simulated	53.17	54.74	58.68	47.88	49.89	55.28
ML-RV / RC, simulated	52.01	53.40	55.38	47.11	48.75	51.45
ML-RV / ML-RC, simulated	52.77	53.80	55.66	47.74	49.19	52.00
Benchmarks						
Unrestricted DCC-GARCH-X	52.95	102.35	100.60	47.92	89.71	90.32
DCC-GARCH-X, horizon-matched	53.64	55.55	56.92	48.28	50.27	52.67
Unrestricted, horizon-matched	57.58	57.84	59.04	51.83	52.67	54.80

QLIKE of the forecast covariance against the mean outer product of the daily returns over the h days (against returns) and against the mean rescaled realized covariance (against the realized covariance), on the origins where every model and both proxies are defined. A model labelled "simulated" averages the inverse of the coordinates over draws of the driver's forecast path at $h > 1$.

Table S4 – Dow Jones panel: one change at a time, against returns

	ΔQLIKE (p value)		
Comparison	$h = 1$	$h = 5$	$h = 22$
ML-RV vs RV, volatility driver			
correlation space, RC	−1.16 (<0.001)	−1.33 (<0.001)	−3.25 (<0.001)
correlation space, ML-RC	−1.19 (<0.001)	−1.34 (<0.001)	−3.48 (<0.001)
coordinates, RC	−1.16 (<0.001)	−1.34 (<0.001)	−3.29 (<0.001)
coordinates, ML-RC	−1.23 (<0.001)	−1.35 (<0.001)	−3.51 (<0.001)
Coordinates vs correlation space			
RV / RC	−0.30 (<0.001)	−0.19 (<0.001)	−0.07 (0.002)
ML-RV / RC	−0.29 (<0.001)	−0.21 (<0.001)	−0.11 (<0.001)
RV / ML-RC	−0.06 (0.172)	−0.09 (0.004)	−0.03 (0.189)
ML-RV / ML-RC	−0.09 (0.018)	−0.11 (0.001)	−0.06 (0.004)
ML-RC vs RC, correlation driver			
correlation space, RV	0.58 (<0.001)	0.31 (<0.001)	0.47 (<0.001)
correlation space, ML-RV	0.55 (<0.001)	0.30 (<0.001)	0.24 (0.006)
coordinates, RV	0.82 (<0.001)	0.41 (<0.001)	0.51 (<0.001)
coordinates, ML-RV	0.75 (<0.001)	0.40 (<0.001)	0.29 (0.004)
Simulated vs plug-in inverse			
RV / RC	0.00 (–)	−0.01 (0.009)	−0.06 (<0.001)
ML-RV / RC	0.00 (–)	−0.02 (0.001)	−0.07 (<0.001)
ML-RV / ML-RC	0.00 (–)	−0.01 (0.355)	−0.09 (0.001)
Coordinates, ML-RV / ML-RC, against			
DCC-GARCH-X	−0.70 (0.003)	−1.14 (<0.001)	−3.07 (<0.001)
Unrestricted DCC-GARCH-X	−0.19 (0.372)	−48.54 (<0.001)	−44.85 (0.001)
DCC-GARCH-X, horizon-matched	−0.87 (<0.001)	−1.74 (<0.001)	−1.17 (<0.001)
Unrestricted, horizon-matched	−4.81 (<0.001)	−4.03 (0.001)	−3.29 (0.030)

Difference in mean QLIKE against returns, second model minus first, with the p value of the Diebold–Mariano test with Newey–West standard errors in parentheses; a negative difference means that the change named first lowers the loss.

Table S5 – Dow Jones panel: one change at a time, against the realized covariance

	ΔQLIKE (p value)		
Comparison	$h = 1$	$h = 5$	$h = 22$
ML-RV vs RV, volatility driver			
correlation space, RC	−0.79 (<0.001)	−1.13 (<0.001)	−3.79 (<0.001)
correlation space, ML-RC	−0.83 (<0.001)	−1.16 (<0.001)	−4.00 (<0.001)
coordinates, RC	−0.77 (<0.001)	−1.14 (<0.001)	−3.82 (<0.001)
coordinates, ML-RC	−0.86 (<0.001)	−1.17 (<0.001)	−4.01 (<0.001)
Coordinates vs correlation space			
RV / RC	−0.25 (<0.001)	−0.18 (<0.001)	−0.07 (0.003)
ML-RV / RC	−0.24 (<0.001)	−0.19 (<0.001)	−0.11 (<0.001)
RV / ML-RC	−0.06 (0.039)	−0.07 (0.014)	0.07 (0.016)
ML-RV / ML-RC	−0.08 (0.042)	−0.08 (0.019)	0.05 (0.040)
ML-RC vs RC, correlation driver			
correlation space, RV	0.53 (<0.001)	0.38 (<0.001)	0.68 (<0.001)
correlation space, ML-RV	0.48 (<0.001)	0.36 (<0.001)	0.47 (<0.001)
coordinates, RV	0.72 (<0.001)	0.50 (<0.001)	0.82 (<0.001)
coordinates, ML-RV	0.64 (<0.001)	0.47 (<0.001)	0.63 (<0.001)
Simulated vs plug-in inverse			
RV / RC	0.00 (–)	−0.01 (<0.001)	−0.07 (<0.001)
ML-RV / RC	0.00 (–)	−0.02 (<0.001)	−0.08 (<0.001)
ML-RV / ML-RC	0.00 (–)	−0.04 (0.001)	−0.16 (<0.001)
Coordinates, ML-RV / ML-RC, against			
DCC-GARCH-X	−0.39 (0.063)	−0.85 (<0.001)	−3.27 (<0.001)
Unrestricted DCC-GARCH-X	−0.18 (0.276)	−40.48 (<0.001)	−38.17 (<0.001)
DCC-GARCH-X, horizon-matched	−0.54 (<0.001)	−1.04 (<0.001)	−0.51 (0.057)
Unrestricted, horizon-matched	−4.08 (<0.001)	−3.44 (0.003)	−2.64 (0.087)

As Table S4, with QLIKE computed against the mean rescaled realized covariance.

Table S6 – Dow Jones panel: the level of the variance forecasts by horizon

		Proxy variance / forecast variance		
Block	Proxy	$h = 1$	$h = 5$	$h = 22$
RV	realized	0.919	0.847	0.658
	returns	0.947	0.873	0.678
ML-RV (linked)	realized	1.030	0.949	0.894
	returns	1.061	0.978	0.921

Mean over stocks and origins of the proxy variance, realized or squared return averaged over the h days, divided by the mean forecast variance.

Table S7 – Dow Jones panel: level corrections and the retransformation

		ΔQLIKE, corrected minus uncorrected (p value)		
Model	Proxy	$h = 1$	$h = 5$	$h = 22$
Variance scale, heavy volatility				
RV / RC (DCC-GARCH-X)	realized	0.18 (0.051)	−0.07 (0.535)	−2.00 (0.024)
	returns	0.20 (0.065)	0.36 (0.037)	−0.90 (0.240)
Variance scale, linked volatility				
ML-RV / RC	realized	−0.01 (0.949)	−0.08 (0.520)	0.94 (0.136)
	returns	0.27 (0.043)	0.46 (0.005)	1.58 (0.005)
Coordinate shift, coordinate				
RV / RC	realized	−0.01 (0.771)	0.01 (0.866)	0.08 (0.198)
	returns	−0.10 (0.012)	−0.08 (0.082)	−0.01 (0.808)
ML-RV / RC	realized	−0.01 (0.847)	−0.01 (0.861)	0.01 (0.909)
	returns	−0.09 (0.008)	−0.09 (0.025)	−0.09 (0.063)
ML-RV / ML-RC	realized	0.04 (0.212)	0.04 (0.358)	−0.20 (0.004)
	returns	−0.16 (0.001)	−0.03 (0.434)	−0.13 (0.024)
Coordinate shift, correlation (placebo)				
RV / RC (DCC-GARCH-X)	realized	−0.00 (0.993)	0.02 (0.783)	0.07 (0.283)
	returns	−0.12 (0.019)	−0.09 (0.098)	−0.02 (0.679)
ML-RV / RC	realized	−0.00 (0.951)	−0.01 (0.853)	−0.01 (0.912)
	returns	−0.11 (0.009)	−0.11 (0.020)	−0.11 (0.040)
ML-RV / ML-RC	realized	0.04 (0.291)	0.01 (0.753)	−0.17 (0.010)
	returns	−0.14 (0.001)	−0.05 (0.202)	−0.13 (0.025)
Simulated retransformation, coordinate				
RV / RC	realized	–	−0.02 (<0.001)	−0.08 (<0.001)
	returns	–	−0.01 (0.011)	−0.07 (<0.001)
ML-RV / RC	realized	–	−0.02 (<0.001)	−0.09 (<0.001)
	returns	–	−0.02 (0.002)	−0.08 (<0.001)
ML-RV / ML-RC	realized	–	−0.04 (0.001)	−0.17 (<0.001)
	returns	–	−0.01 (0.431)	−0.10 (0.001)

Difference in mean QLIKE, corrected minus uncorrected forecast, against the realized covariance and against returns, with Diebold–Mariano p values. The variance scale and the coordinate shift use only the 250 origins whose targets are known at the origin; the shift on correlation-space models is a placebo, since they involve no retransformation; the simulated retransformation replaces the plug-in inverse at $h > 1$.

Table S8 – Dow Jones panel: losses against the realized covariance, all origins

	QLIKE			Frobenius		
Model	$h = 1$	$h = 5$	$h = 22$	$h = 1$	$h = 5$	$h = 22$
Correlation space						
RV / RC (DCC-GARCH-X)	48.132	50.083	55.427	21771.4	11298.5	11508.5
ML-RV / RC	47.343	48.950	51.633	22742.4	11834.4	10207.5
RV / ML-RC	48.660	50.465	56.104	22357.5	12472.3	12025.4
ML-RV / ML-RC	47.825	49.310	52.104	23345.9	12788.3	10624.5
Coordinates						
RV / RC	47.881	49.902	55.352	21660.4	11222.3	11625.0
ML-RV / RC	47.106	48.763	51.528	22395.5	11632.8	10229.1
RV / ML-RC	48.598	50.398	56.172	22284.1	12684.0	12311.3
ML-RV / ML-RC	47.742	49.230	52.158	22871.6	12668.8	10691.1
RV / RC, simulated	47.881	49.888	55.281	21660.4	11239.5	11548.1
ML-RV / RC, simulated	47.106	48.746	51.446	22395.5	11681.8	10268.4
ML-RV / ML-RC, simulated	47.742	49.189	52.000	22871.6	12742.4	10762.7
Benchmarks						
Unrestricted DCC-GARCH-X	47.923	89.715	90.324	21664.7	$> 10^6$	$> 10^6$
DCC-GARCH-X, horizon-matched	48.277	50.267	52.672	25458.6	13812.6	10138.8
Unrestricted, horizon-matched	51.826	52.670	54.800	$> 10^6$	$> 10^6$	$> 10^6$

QLIKE and Frobenius distance of the forecast covariance against the mean rescaled realized covariance over the h days, averaged over all forecast origins of each model. A model labelled "simulated" averages the inverse of the coordinates over draws of the driver's forecast path at $h > 1$.

Table S9 – Dow Jones panel: minimum variance portfolios

Forecast	Return (%)	Realized risk (%)	Return s.d. (%)	Sharpe	Turnover	Gross exposure
$h = 1$, 2,352 holding periods						
DCC-GARCH-X	8.68	10.58	13.74	0.49	1.11	1.77
DCC-GARCH-X, horizon-matched	10.87	11.08	14.06	0.63	1.18	1.80
Unrestricted DCC-GARCH-X	8.09	10.75	13.71	0.44	1.20	1.76
Unrestricted, horizon-matched	8.09	10.94	15.41	0.39	0.51	1.57
Coordinates, RV / RC	8.11	10.52	13.73	0.44	1.12	1.77
Coordinates, ML-RV / RC	8.72	10.42	13.58	0.49	0.86	1.67
$h = 5$, 470 holding periods						
DCC-GARCH-X	9.80	11.06	12.17	0.64	1.36	1.77
DCC-GARCH-X, horizon-matched	11.08	11.49	12.86	0.71	1.53	1.78
Unrestricted DCC-GARCH-X	−0.08	16.97	23.35	−0.09	0.64	1.59
Unrestricted, horizon-matched	7.28	11.25	13.45	0.39	0.72	1.53
Coordinates, RV / RC	9.39	11.02	12.15	0.61	1.36	1.77
Coordinates, ML-RV / RC	9.86	10.81	11.81	0.67	0.95	1.61
$h = 22$, 106 holding periods						
DCC-GARCH-X	6.39	11.97	12.38	0.35	0.99	1.85
DCC-GARCH-X, horizon-matched	9.92	11.77	11.26	0.70	1.36	1.64
Unrestricted DCC-GARCH-X	2.58	17.38	21.23	0.03	0.62	1.51
Unrestricted, horizon-matched	6.66	11.68	11.13	0.42	0.71	1.44
Coordinates, RV / RC	6.11	11.95	12.42	0.33	0.98	1.85
Coordinates, ML-RV / RC	9.26	11.33	10.67	0.68	0.98	1.57

Realized risk is the annualized square root of the portfolio variance under the mean rescaled realized covariance of the holding period, and return s.d. the annualized standard deviation of the realized portfolio return; return and Sharpe ratio are annualized, with a risk-free rate of 2%. Turnover is the sum of the absolute changes in the weights between consecutive holding periods, and gross exposure the sum of the absolute weights. The forecasts of the unrestricted DCC-GARCH-X at $h > 1$ and of its horizon-matched version are near-singular at some origins, and their portfolios are not interpretable.

Table S10 – Dow Jones panel: differences and fees against DCC-GARCH-X

Forecast	Δ realized risk (t)	Δ return s.d. (t)	Fee, $\gamma = 5$	Fee, $\gamma = 10$	Paired t	Fee, $\gamma = 5$, without largest
$h = 1$, 2,352 holding periods						
DCC-GARCH-X, horizon-matched	0.50 (8.20)	0.32 (2.23)	1.97	1.75	1.55	1.81
Unrestricted DCC-GARCH-X	0.17 (3.95)	−0.03 (−0.17)	−0.56	−0.54	−0.31	−1.05
Unrestricted, horizon-matched	0.36 (1.96)	1.67 (1.26)	−1.81	−3.03	−0.21	−2.56
Coordinates, RV / RC	−0.06 (−5.11)	−0.01 (−0.25)	−0.56	−0.55	−1.56	−0.64
Coordinates, ML-RV / RC	−0.16 (−1.84)	−0.16 (−0.74)	0.15	0.26	0.03	−0.31
$h = 5$, 470 holding periods						
DCC-GARCH-X, horizon-matched	0.43 (4.72)	0.69 (2.35)	0.84	0.40	0.97	1.34
Unrestricted DCC-GARCH-X	5.91 (4.92)	11.18 (2.43)	−19.94	−30.52	−1.41	−16.39
Unrestricted, horizon-matched	0.19 (0.70)	1.27 (1.01)	−3.35	−4.19	−0.79	−1.46
Coordinates, RV / RC	−0.04 (−2.91)	−0.03 (−0.71)	−0.40	−0.38	−1.39	−0.45
Coordinates, ML-RV / RC	−0.25 (−2.54)	−0.36 (−1.53)	0.28	0.50	0.04	−0.60
$h = 22$, 106 holding periods						
DCC-GARCH-X, horizon-matched	−0.21 (−1.70)	−1.12 (−1.96)	4.21	4.92	2.07	3.39
Unrestricted DCC-GARCH-X	5.40 (2.99)	8.85 (1.34)	−11.52	−20.73	−0.62	−11.99
Unrestricted, horizon-matched	−0.29 (−1.19)	−1.25 (−1.73)	1.02	1.80	0.13	1.93
Coordinates, RV / RC	−0.02 (−1.81)	0.04 (0.37)	−0.30	−0.33	−1.52	−0.24
Coordinates, ML-RV / RC	−0.65 (−7.60)	−1.71 (−2.25)	3.87	4.92	1.55	4.16

Differences to DCC-GARCH-X in realized risk and in return s.d., in percentage points, with the Newey–West t statistic of the difference in realized risk and in squared portfolio returns in parentheses, with $\lfloor n^{1/3} \rfloor$ lags, 13, 7 and 4 at $h = 1$, 5 and 22. Annual fee, in percent, that an investor with relative risk aversion γ would pay to switch from DCC-GARCH-X to the forecast (Fleming et al., 2003), with the paired t statistic of the return differences; a negative fee means the investor would need to be paid, and without largest drops the holding period with the largest absolute return difference. The rows of the near-singular forecasts are not interpretable (Table S9).

S4 S&P 500 panel

S4.1 The volatility block

On the monthly panel the learner, rather than the recursion, improves the volatility forecast. Table S11 shows that over the 183 months QLIKE against the rescaled realized variance is 0.313 for the linked block and 0.329 for the RV block ($t = -1.38$), and 0.267 for the learner alone, rescaled to the level of returns ($t = -2.08$). On the returns likelihood the learner alone is also the lowest ($t = -2.22$), while the linked block does not differ significantly from the RV block ($t = -1.46$). However, the learner alone is not calibrated to returns: the mean squared daily return is 1.183 times its forecast variance, against 1.007 for the linked block and 1.045 for the RV block. For this reason the correlation block standardizes returns with the linked block. With 36-month windows the learner alone also attains the lowest QLIKE, 0.277 against 0.288 for the linked block and 0.299 for the RV block, and on the fifteen-minute grid the order of the 60-month windows holds, with 0.272, 0.284 and 0.243.

S4.2 All forecasts, both grids

Tables S12 and S13 report every forecast on the two grids, including the links in correlation space (rows starting with c), the pairwise features of Bollerslev et al. (2026) (BLT) as a learned driver in coordinates, and DCC-NL. Against the returns every link attains a lower likelihood than HAR and the corrected projections of Bollerslev et al. (2026). On the five-minute grid the difference to the corrected LASSO is between -10.9 and -11.8 per day (Newey–West t between -3.27 and -2.96), and it does not rest on a few months: at the median it is between -10.1 and -11.1, without the largest month between -7.6 and -8.6, and each link is better in 68% to 70% of months (Table S20). The projections without the correction have a likelihood only in the months in which they are positive definite. In those months SHAR-F-Exp is worse than the shrunk target by 32.0 ($t = 1.92$, 30 months), and the LASSO by 9.5 ($t = 3.87$, 133 months). Every link improves on the target, by 0.42 to 1.30 per day (t between -6.25 and -3.47), with a negative median difference for every link and in 67% to 90% of months. The gain holds in both periods: before January 2021 the links attain 113.78 to 114.38 against 114.67 for the target, and over the 64 months since then 95.79 to 97.61, against 98.27 for the target and 103.63 for the corrected LASSO.

DCC-NL is worse than the target by 31.5 per day ($t = 6.00$) and better in 9% of months. Its correlations are estimated on returns standardized by an exponentially weighted variance, and their mean of 0.293 is below the 0.351 of the target, which is estimated on the returns standardized by the volatility block on which every forecast is scored. We therefore do not read this comparison as evidence against DCC-NL itself.

Against the rescaled realized measures, QLIKE ranks the forecasts as the likelihood does, while the squared-error criteria rank them in the reverse order. The corrected LASSO attains a QLIKE of 213.7 on the realized correlation against 168.6 to 170.1 for the links (t between -24.28 and -23.97), and 303.8 on the realized covariance against 249.7 to 250.7 (t between -7.63 and -7.40). Its Frobenius distance is 45.4 against 51.8 to 52.7 for the links (t between 3.96 and 4.42), and its pooled R^2 is 8.8% against -17% to -21% for the links, whose R^2 is -15% to -18% on the fifteen-minute grid. The two squared errors are computed pair by pair, the way the projections are fitted, while QLIKE, like the likelihood, weighs a forecast through its inverse. The rescaling narrows the disagreement about the level: the rescaled realized correlation averages 0.283 over the evaluation months, BLT's projections of it 0.30 and the links 0.35. Before rescaling, the realized correlation averaged 0.254 and the projections 0.26, and every realized criterion reversed the order of the likelihood: the corrected LASSO attained a QLIKE of 159.8 on the realized correlation against 190.3 to 191.0 for the links, and a pooled R^2 of 8.1% against -66% to -70%.

On the fifteen-minute grid every link attains a lower likelihood than the corrected LASSO, by 12.2 to 13.1 per day (Newey–West t between -2.69 and -2.49), and for the correlation-space link on the lagged realized correlation the difference is -9.9 at the median and -7.5 without the largest month, lower in 69% of months (Table S21). QLIKE again ranks the forecasts as the likelihood does, 295.2 for the corrected LASSO on the correlation against 256.7 to 257.9 for the links, while the pooled R^2 is 10.0% for the corrected LASSO and between -15% and -18% for the links. SHAR-F-Exp is indefinite in 63.4% of the 183 months and the LASSO in 20.8%, while no link or target forecast is.

S4.3 Correlation space and the learned driver

With the drivers moving only the structure of the correlations, running the recursion in coordinates lowers the returns likelihood by 0.32 with the lagged realized correlation ($t = -3.00$, better in 80% of months) and by 0.33 with the learned driver ($t = -1.21$), and replacing the lagged realized correlation by the learned driver lowers it by 0.46 in correlation space ($t = -3.41$) and by 0.46 in coordinates ($t = -2.95$). The fifteen-minute grid gives the same ranking: the learned driver lowers the likelihood by 0.30 in correlation space and by 0.66 in coordinates ($t = -2.67$ and -4.76). The projection of Bollerslev et al. (2026) does not improve on the lagged realized correlation as a driver on either grid, with differences of $+0.09$ and $+0.07$ ($t = 1.67$ and 1.00). Against the rescaled covariance the learned driver also lowers the risk of the minimum variance portfolio, by 0.06 percentage points in correlation space ($t = -6.37$) and by 0.06 in coordinates ($t = -4.41$), and against the rescaled realized correlation the coordinate recursion on the lagged realized correlation is better by 0.63 in QLIKE ($t = -5.43$).

The link that uses the LASSO projection of Bollerslev et al. (2026) as its driver, with their fitted values, is better than their corrected LASSO by 10.9 per day ($t = -2.97$) on the five-minute grid and by 12.2 on the fifteen-minute grid ($t = -2.51$).

S4.4 The weights of the links

The likelihood puts little weight on the lagged realized correlation and on the projection of Bollerslev et al. (2026), and more on the learned drivers (Tables S14 and S15). On the five-minute grid the mean weight is 0.015 on the lagged realized correlation in correlation space, below 0.05 at 94% of origins, 0.025 in coordinates, 0.035 on BLT's corrected LASSO and 0.064 on BLT's pairwise features in coordinates. The learned driver receives 0.104 in correlation space and 0.164 in coordinates, below 0.05 at only 9% and 18% of origins, with a persistence of 0.13 and 0.14, so the structure it supplies enters mainly as a new signal rather than through the state, and its weights rise after January 2021 to 0.13 and 0.27. On the fifteen-minute grid the weights are 0.012 and 0.015 on the lagged realized correlation, 0.029 on BLT's projection, and 0.080 and 0.133 on the learned driver, which rise to 0.108 and 0.223 after January 2021. The small weights on the lagged realized correlation and on the projection are the pattern of Section S2.2, where, for the projection in the noisy cells with $n = 27$, the QML weight fell from

0.63 at $k = 30$ to 0.20 at $k = 300$, while BLT's criterion selected between 0.77 and 0.96 in the noisy cells at every k.

S4.5 Economic value

How the minimum variance portfolios rank depends on how their risk is measured (Tables S17 and S18). Against the rescaled realized covariance every link builds a less risky portfolio than the projections of Bollerslev et al. (2026): on the five-minute grid 8.15% to 8.23% a year, against 10.15% for the corrected LASSO and 9.93% for HAR, with 8.29% for the shrunk target and 8.57% for DCC-NL. The difference to the corrected LASSO is between -1.93 and -2.01 percentage points (t between -11.98 and -11.49), and every link lowers the risk of the target's portfolio, by 0.06 to 0.14 points (t between -7.09 and -4.55). However, the rescaled covariance takes its average from the shrunk second moment of returns, which is also what the target of the links estimates, and the ranking against the projections does not hold in the returns of the portfolios. The annualized standard deviation of the realized monthly return is 11.88% to 11.95% for the links and 11.96% for the target, against 11.62% for the corrected LASSO, 11.01% for DCC-NL and 10.75% for HAR, and no difference in the squared monthly returns is significant, for the links against the corrected LASSO (t between 0.51 and 0.67) or for HAR against it ($t = -1.08$). Before rescaling, the realized covariance also placed HAR (10.37%) and the corrected LASSO (10.48%) below the links (10.67% to 10.74%). An investor with a relative risk aversion of five would need to be paid between 0.3% and 0.5% a year to switch from the corrected LASSO to a link (Fleming et al., 2003), but the paired t statistics of the return differences are below 0.2 in absolute value (Table S19). On the fifteen-minute grid the links' portfolios are less risky than the corrected LASSO's against the rescaled covariance by 1.64 to 1.72 points (t between -10.93 and -10.33), the standard deviations of their realized returns, 11.68% to 11.75% against 11.52%, do not differ significantly ($t \leq 0.71$), and their turnover is 1.95 to 2.02 against 2.95. The simulation of Section S2.2 showed the same pattern: at $k = 300$ the projection attained a lower risk of the minimum variance portfolio than the QML-selected link, 13.69 against 14.26, with a higher true QLIKE, 35.4 against 24.2.

S4.6 Variants of the correlation block

Level of the driver kept. Tables S22 and S23 also report the links in which the driver deviations enter with their mean over the pairs, that is, the recursions of Section 2.3 with $\mathcal{S}$ the identity. We fixed the rule for choosing between the two forms before computing the structure-only results: the structure-only form is adopted if it gives the learned coordinates a lower returns likelihood on the five-minute grid and the fifteen-minute grid agrees in sign. The differences are -1.36 per day on the five-minute grid ($t = -1.71$) and -0.32 on the fifteen-minute grid ($t = -0.61$). Neither is significant, and the second rests on March 2020: without that month it is $+0.09$, and its median is $+0.37$. The comparisons within each form are sharper, because the links of one form share their level. With the level kept, every link remains better than the corrected LASSO, by 10.5 to 11.6 per day (t between -3.99 and -3.06), but the drivers change the likelihood of the target by between -1.07 and $+0.05$, only the difference of the correlation-space link on the lagged realized correlation is significant at the 5% level (-0.55, $t = -2.16$), and the learned coordinates are worse than the target by 0.05, against better by 1.30 ($t = -5.55$) with the structure only. The weight on the learned coordinates is then 0.077, with a persistence of 0.50. With the level kept, the learned coordinates deviated from their mean level by only 0.002 to 0.003 between December 2012 and March 2013, and at a persistence of 0.92 the mean forecast correlation fell 0.04 to 0.06 below that of the target.

Drivers from the rescaled measure. When the structure-only drivers are built from the rescaled measure, the likelihood puts a mean weight of at most 0.001 on the lagged realized correlation and on the learned coordinates, below 0.05 at every origin on both grids, so these four links reduce to the shrunk target, and they are worse than the links on the measure before rescaling by 0.52 to 1.30 per day on the five-minute grid and by 0.51 to 1.31 on the fifteen-minute grid (t between 4.01 and 6.44). Only the links on BLT's projection and on the pairwise features keep their weights, 0.037 and 0.063 on the five-minute grid. The rescaling does not remove the structure of the measure pair by pair. Month by month, the correlation across pairs between the rescaled and the original realized correlation, each less its mean over the pairs, is 0.93 on the five-minute grid and 0.95 on the fifteen-minute grid, and their deviations from their means are equally correlated across pairs with the next month's sample correlation of returns less that of the preceding 48 months, 0.139 against 0.141. It removes what the likelihood uses.

Added to the target with a weight of 0.01, the deviation of the lagged realized correlation lowers the returns likelihood of the next month by 0.29 per day ($t = -4.21$), and the objective of the estimation window at 74% of origins, while the rescaled deviation changes the next month's likelihood by +0.01 ($t = 0.10$) and lowers the objective at 0.5% of origins; on the fifteen-minute grid the changes are -0.32 ($t = -4.97$) and -0.04 ($t = -0.63$). A correlation across pairs weighs every pair equally, while the likelihood weighs a deviation through the inverse of the target, most in the directions in which the target has small eigenvalues.

Learner of the coordinates. Bollerslev et al. (2026) fit their projections with the LASSO, while our learned coordinates are boosted. With the LASSO as the learner on the same features, the likelihood puts a weight of 0.118 on the learned coordinates on the five-minute grid and 0.081 on the fifteen-minute grid, and the link improves on the target by 0.93 and 0.79 per day ($t = -6.23$ and -5.64), against 1.30 and 1.31 with boosting. Boosting is better in coordinates, by 0.37 and 0.53 ($t = -2.73$ and -3.90), and in correlation space the two learners do not differ significantly ($t = -1.37$ and 0.26). With the LASSO the learned coordinates remain better than the corrected LASSO by 11.4 per day ($t = -3.15$), but they no longer improve significantly on the lagged realized correlation ($t = -1.02$ and -1.79).

Shorter windows. With 36-month windows, for the volatility block, the target and the links, the likelihood puts a mean weight of 0.000 on the lagged realized correlation and on BLT's projection, below 0.05 at every origin, so these links are nearly identical to the shrunk target, and a weight of 0.016 on the learned coordinates, which then improve on the target by 0.14 ($t = -2.21$). Every forecast of this variant remains better than the corrected LASSO, by 12.2 to 12.4 per day (t between -3.24 and -3.20), and against the rescaled covariance its minimum variance portfolios are less risky than the corrected LASSO's, 8.49% against 10.22% a year ($t = -9.13$).

RV volatility. When the returns are standardized by the RV volatilities instead of the linked block, the advantage of the links over the corrected LASSO shrinks: on the five-minute grid it is 7.7 to 8.5 per day (t between -2.33 and -2.07), while the shrunk target alone is better by 7.2 ($t = -1.95$), and on the fifteen-minute grid it is 9.0 to 9.9 (t between -2.00 and -1.82), with 8.6 for the target ($t = -1.72$). Every link improves on the shrunk target of the same scoring, by 0.46 to 1.24 on the five-minute grid (t between -6.64 and -4.64) and by 0.48 to 1.35 on the

fifteen-minute grid (t between -6.50 and -5.23).

Level form. When the driver enters the recursion in levels rather than in deviations from its mean, the likelihood puts a mean weight of at most 0.002 on it, below 0.05 at every origin, so these links collapse to the shrunk target: their likelihood is within 0.33 of the target's and no difference is significant. This is consistent with the level bias of Section 3.2: in levels the driver pulls the forecast towards the level of the realized correlation, 0.25, while the returns have a correlation of 0.35. The links remain better than the corrected LASSO, by 10.5 to 10.8 per day (t between -2.95 and -2.87).

Sample target. Replacing the nonlinearly shrunk target by the sample correlation of the same 48 months of standardized returns makes the links worse than the shrunk target by 99.3 to 105.2 per day (t between 6.75 and 7.28) and worse than the corrected LASSO by 88.8 to 94.7 (t between 6.52 and 6.93), with smallest eigenvalues between 0.002 and 0.023. At 315 stocks the advantage of the links rests on the shrinkage of the target, as Section 2.4 anticipates.

S4.7 The validity of the projections

Fitted to the measure before rescaling, the projections of Bollerslev et al. (2026) are indefinite in 58.5% and 13.7% of months on the five-minute grid and in 45.4% and 10.4% on the fifteen-minute grid; fitted to the rescaled measure, in 83.6% and 27.3%, and in 63.4% and 20.8%. On the five-minute grid a smallest eigenvalue below 0.05 adds 7.1 and 8.8 percentage points, and on the fifteen-minute grid 9.8 and 8.7.

Whether a projection is valid is largely decided when it is fitted. The projections are refitted once a year, and their failures cluster by refit year: on the five-minute grid SHAR-F-Exp is indefinite in every month of 2018 to 2021 and of 2023 to 2025 and in 17% of those of 2022, and the LASSO in 92% of the months of 2020 and in none of those of 2011, 2012, 2014, 2016 and 2019. Dummies for the refit year explain between 31% and 50% of the variance of the monthly failure indicator for the two projections on the two grids, against a 95th percentile of at most 14% when the months are permuted, while the correlation of the failure indicator with the month's mean realized correlation lies between -0.06 and 0.06.

A forecast that is positive definite can still be too close to singular for the returns likelihood. Table S24 groups the months by the smallest eigenvalue of the uncorrected projection. In the

two months in which that eigenvalue is below 0.01 for the LASSO, its likelihood is worse than that of the shrunk target by 74.1 per day, and in the 14 months between 0.01 and 0.05 by 20.4, against 7.2 in the 104 months above 0.10, while the correlation-space link on the lagged realized correlation is better than the target by between 0.23 and 0.91 in the same months. The convexity correction in the replication code of Bollerslev et al. (2026), which mixes a forecast with their HAR projection until its smallest eigenvalue reaches 0.1, removes most of this cost. Table S25 rescores every forecast with triggers of 0.02, 0.05, 0.1 and 0.2. The likelihood of the corrected LASSO is 127.20, 122.02, 119.45 and 122.09 at these triggers, and the correlation-space link on the lagged realized correlation remains better at every trigger, by 11.0 to 18.8 per day (t between -3.92 and -2.96), as do the learned coordinates (t between -4.23 and -3.27). Against the rescaled covariance the minimum variance portfolio of this link is also less risky than that of the corrected LASSO at every trigger, by 1.41 to 2.81 percentage points (t between -11.78 and -9.16).

Table S11 – S&P 500 panel, five-minute grid: the volatility block

Volatility block	Months	QLIKE vs RV	t against RV	Returns likelihood	t against RV	Squared return / forecast
ML-RV (linked)	183	0.313	−1.38	−145.686	−1.46	1.007
RV	183	0.329	–	−145.299	–	1.045
ML-RV alone, rescaled	183	0.267	−2.08	−146.681	−2.22	1.183

Means over the evaluation months. ML-RV (linked) is the two-step block with the learned variance as its driver, RV the same recursion on the lagged realized variance, and ML-RV alone the learned variance rescaled to the level of returns. QLIKE is computed against the month's rescaled realized variance per day; the returns likelihood is the Gaussian quasi-likelihood of the month's daily returns, per stock; t is the Newey–West t statistic of the monthly difference against RV; the last column is the mean squared daily return divided by the forecast.

Table S12 – S&P 500 panel, five-minute grid: all forecasts

Forecast	Returns likelihood	QLIKE, covariance	QLIKE, correlation	Frobenius	R^2 vs HAR	Indefinite (%)	Smallest eigenvalue
Shrunk target	108.94	251.01	170.44	52.92	−21.89	0.0	0.072
Lagged RC	403.03	485.68	356.80	47.61	−3.37	0.0	0.020
HAR	134.77	315.95	228.09	47.87	0.00	0.0	0.147
SHAR-F-Exp	140.38	305.87	244.83	45.88	5.93	83.6	−1.511
SHAR-F-Exp, corrected	126.89	311.53	224.49	45.79	7.39	0.0	0.100
LASSO	99.10	277.57	211.81	45.52	7.60	27.3	−0.664
LASSO, corrected	119.45	303.84	213.68	45.38	8.82	0.0	0.100
DCC-NL	140.44	290.70	213.60	45.86	6.27	0.0	0.047
c, RC	108.42	250.55	170.06	52.73	−21.07	0.0	0.068
c, BLT LASSO	108.51	250.67	170.05	52.72	−21.05	0.0	0.068
c, ML-RC	107.96	249.88	169.49	52.42	−19.83	0.0	0.061
g, RC	108.10	250.20	169.43	52.48	−20.02	0.0	0.071
g, ML-RC	107.64	249.75	168.64	51.82	−17.14	0.0	0.069
g, ML-RC, pairwise	107.79	249.61	169.08	52.13	−18.37	0.0	0.063

A row starting with c is a link in correlation space and one starting with g a link in coordinates, followed by its correlation driver: RC for the lagged realized correlation, ML-RC for the learned coordinates, BLT LASSO for the corrected LASSO projection, and pairwise for BLT's pairwise features fitted to the coordinates. Every link uses the ML-RV volatility block. Other definitions as in Table S13.

Table S13 – S&P 500 panel, fifteen-minute grid of Bollerslev et al. (2026): all forecasts

Forecast	Returns likelihood	QLIKE, covariance	QLIKE, correlation	Frobenius	R^2 vs HAR	Indefinite (%)	Smallest eigenvalue
Shrunk target	115.33	337.09	258.25	53.84	−18.97	0.0	0.073
Lagged RC	824.46	972.94	834.03	50.99	−9.13	0.0	0.020
HAR	149.29	401.65	317.10	49.39	0.00	0.0	0.206
SHAR-F-Exp	95.52	385.10	319.63	46.78	8.56	63.4	−1.659
SHAR-F-Exp, corrected	134.30	392.57	306.52	46.81	9.08	0.0	0.100
LASSO	147.97	400.39	300.70	46.57	9.53	20.8	−0.710
LASSO, corrected	127.07	384.06	295.19	46.51	10.04	0.0	0.100
DCC-NL	147.15	376.04	298.81	47.02	8.02	0.0	0.047
DCC-NL, monthly	133.14	353.36	269.55	52.63	−14.21	0.0	0.052
c, RC	114.82	336.65	257.86	53.68	−18.32	0.0	0.071
c, BLT LASSO	114.89	336.74	257.81	53.68	−18.32	0.0	0.071
c, ML-RC	114.52	336.28	257.34	53.39	−17.15	0.0	0.067
g, RC	114.68	336.46	257.54	53.56	−17.81	0.0	0.072
g, ML-RC	114.01	335.98	256.73	52.90	−15.01	0.0	0.071
g, ML-RC, pairwise	114.17	335.69	256.81	53.01	−15.47	0.0	0.063

Realized measures from fifteen-minute returns, 183 months from February 2011 to April 2026. Row labels as in Table S12; HAR, SHAR-F-Exp and LASSO are the projections of Bollerslev et al. (2026), fitted to the rescaled measure, with their convexity correction where marked. The returns likelihood is per day; QLIKE and Frobenius are against the rescaled realized covariance and correlation; R^2 is the pooled out-of-sample R^2 of Bollerslev et al. (2026) against their HAR projection; indefinite is the share of months in which the forecast is not positive definite.

Table S14 – S&P 500 panel, five-minute grid: the weights of the links

Arm	Origins	Mean α	Median α	Share $\alpha < 0.05$	Mean β
main: c blt	183	0.035	0.031	0.78	0.208
main: c learn	183	0.104	0.093	0.09	0.128
main: c rc	183	0.015	0.012	0.94	0.457
main: g learn	183	0.164	0.161	0.17	0.138
main: g pair	183	0.064	0.052	0.48	0.386
main: g rc	183	0.025	0.022	0.89	0.599

Mean and median weight α on the driver over the 183 origins, the share of origins with $\alpha < 0.05$ and the mean persistence β; c denotes correlation space and g coordinates.

Table S15 – S&P 500 panel, fifteen-minute grid: the weights of the links

Arm	Origins	Mean α	Median α	Share $\alpha < 0.05$	Mean β
main: c blt	183	0.029	0.026	0.81	0.100
main: c learn	183	0.080	0.069	0.27	0.257
main: c rc	183	0.012	0.009	0.99	0.423
main: g learn	183	0.133	0.125	0.19	0.185
main: g pair	183	0.068	0.058	0.39	0.358
main: g rc	183	0.015	0.013	1.00	0.541
dccnlm: dccnl	183	0.007	0.003	1.00	0.773

As Table S14.

Table S16 – S&P 500 panel, five-minute grid: differences against BLT's corrected LASSO

	Mean difference against LASSO, corrected (t)				
Forecast	Returns likelihood	QLIKE, covariance	QLIKE, correlation	Frobenius	GMV risk
Shrunk target	−10.51 (−2.88)	−52.83 (−7.47)	−43.24 (−24.20)	7.54 (4.53)	−1.87 (−11.29)
Lagged RC	283.58 (5.06)	181.84 (3.09)	143.12 (26.81)	2.22 (2.67)	−1.05 (−10.78)
HAR	15.32 (9.16)	12.11 (6.86)	14.41 (11.04)	2.48 (6.00)	−0.23 (−2.76)
SHAR-F-Exp	29.78 (2.32)	35.36 (2.42)	34.65 (2.06)	0.49 (1.74)	0.20 (3.76)
SHAR-F-Exp, corrected	7.43 (5.22)	7.69 (4.00)	10.81 (8.43)	0.41 (2.22)	0.20 (3.76)
LASSO	2.85 (3.15)	2.76 (3.38)	3.41 (2.91)	0.14 (0.77)	0.00 (–)
DCC-NL	20.99 (4.14)	−13.14 (−2.87)	−0.08 (−0.02)	0.48 (0.62)	−1.58 (−10.92)
c, RC	−11.03 (−2.96)	−53.29 (−7.42)	−43.62 (−24.26)	7.34 (4.42)	−1.93 (−11.49)
c, BLT LASSO	−10.94 (−2.97)	−53.17 (−7.47)	−43.63 (−24.25)	7.34 (4.42)	−1.93 (−11.51)
c, ML-RC	−11.49 (−3.02)	−53.96 (−7.40)	−44.19 (−24.28)	7.04 (4.24)	−1.99 (−11.73)
g, RC	−11.36 (−3.09)	−53.64 (−7.52)	−44.25 (−23.97)	7.10 (4.30)	−1.95 (−11.67)
g, ML-RC	−11.82 (−3.27)	−54.09 (−7.63)	−45.04 (−24.20)	6.44 (3.96)	−2.01 (−11.98)
g, ML-RC, pairwise	−11.66 (−3.12)	−54.23 (−7.42)	−44.60 (−23.99)	6.75 (4.17)	−1.98 (−11.78)

Mean monthly loss difference, forecast minus BLT's corrected LASSO, with the Newey–West t statistic in parentheses; a negative difference favours the forecast. QLIKE and Frobenius are computed against the rescaled realized measure, and GMV risk, in percentage points, under the rescaled realized covariance.

Table S17 – S&P 500 panel, five-minute grid: minimum variance portfolios

Forecast	Return (%)	Realized risk (%)	Return s.d. (%)	Sharpe	Turnover	Corrected (%)
Shrunk target	12.34	8.29	11.96	0.86	1.67	26.8
Lagged RC	13.02	9.10	11.26	0.98	4.15	100.0
HAR	13.51	9.92	10.74	1.07	3.07	0.0
SHAR-F-Exp	13.33	10.35	11.51	0.98	3.36	94.5
SHAR-F-Exp, corrected	13.33	10.35	11.51	0.98	3.36	0.0
LASSO	12.65	10.15	11.62	0.92	2.93	43.2
LASSO, corrected	12.65	10.15	11.62	0.92	2.93	0.0
DCC-NL	13.13	8.57	11.01	1.01	2.43	39.3
c, RC	12.35	8.23	11.88	0.87	1.71	32.2
c, BLT LASSO	12.41	8.23	11.92	0.87	1.69	31.7
c, ML-RC	12.44	8.17	11.89	0.88	1.74	31.7
g, RC	12.48	8.21	11.95	0.88	1.70	27.3
g, ML-RC	12.59	8.15	11.95	0.89	1.75	32.2
g, ML-RC, pairwise	12.52	8.17	11.90	0.88	1.70	33.3

Realized risk is the annualized square root of the portfolio variance under the month's rescaled realized covariance, and return s.d. the annualized standard deviation of the realized monthly portfolio return, which uses no realized measure; corrected is the share of months in which the convexity correction of Bollerslev et al. (2026) is applied before the portfolio is formed.

Table S18 – S&P 500 panel, fifteen-minute grid: minimum variance portfolios

Forecast	Return (%)	Realized risk (%)	Return s.d. (%)	Sharpe	Turnover	Corrected (%)
Shrunk target	12.60	8.51	11.76	0.90	1.94	26.2
Lagged RC	12.38	9.59	11.54	0.90	4.74	100.0
HAR	12.44	10.10	10.61	0.98	3.33	0.0
SHAR-F-Exp	12.29	10.34	11.39	0.90	3.49	89.1
SHAR-F-Exp, corrected	12.29	10.34	11.39	0.90	3.49	0.0
LASSO	12.00	10.10	11.52	0.87	2.95	37.2
LASSO, corrected	12.00	10.10	11.52	0.87	2.95	0.0
DCC-NL	13.29	8.70	10.98	1.03	2.56	39.3
DCC-NL, monthly	12.23	8.42	11.40	0.90	2.27	48.6
c, RC	12.52	8.46	11.69	0.90	1.97	31.1
c, BLT LASSO	12.57	8.45	11.73	0.90	1.95	31.1
c, ML-RC	12.58	8.40	11.68	0.91	2.00	31.7
g, RC	12.63	8.46	11.75	0.90	1.96	26.2
g, ML-RC	12.67	8.38	11.70	0.91	2.02	29.0
g, ML-RC, pairwise	12.80	8.39	11.75	0.92	1.97	33.3

As Table S17, on the fifteen-minute grid.

Table S19 – S&P 500 panel, five-minute grid: fees for switching from BLT's corrected LASSO

Forecast	Months	Fee, $\gamma = 5$	Fee, $\gamma = 10$	Paired t	Fee, $\gamma = 5$, without the largest month
Shrunk target	183	−0.5	−0.8	−0.20	−0.8
Lagged RC	183	0.6	0.8	0.33	−0.0
HAR	183	1.4	1.9	0.82	0.2
SHAR-F-Exp	183	0.7	0.8	1.15	1.0
SHAR-F-Exp, corrected	183	0.7	0.8	1.15	1.0
LASSO	183	0.0	0.0	–	0.0
DCC-NL	183	0.8	1.2	0.30	−0.0
c, RC	183	−0.5	−0.6	−0.19	−0.1
c, BLT LASSO	183	−0.4	−0.6	−0.15	−0.1
c, ML-RC	183	−0.4	−0.6	−0.13	−0.0
g, RC	183	−0.4	−0.6	−0.11	−0.0
g, ML-RC	183	−0.3	−0.5	−0.04	−0.0
g, ML-RC, pairwise	183	−0.3	−0.5	−0.08	0.1

Annual fee, in percent, that an investor with relative risk aversion γ would pay to switch from the corrected LASSO to the forecast (Fleming et al., 2003), with the paired t statistic of the monthly return differences; a negative fee means the investor would need to be paid.

Table S20 – S&P 500 panel, five-minute grid: how much single months drive the likelihood comparisons

		Returns likelihood difference			
Forecast	Months	Mean	Median	Mean without largest month	Better (% of months)
Against the shrunk target					
Lagged RC	183	294.10	206.78	239.75	0
HAR	183	25.83	24.89	21.99	11
SHAR-F-Exp	30	31.99	15.88	19.25	23
SHAR-F-Exp, corrected	183	17.95	17.00	15.60	20
LASSO	133	9.52	10.79	8.97	29
LASSO, corrected	183	10.51	10.03	7.24	33
DCC-NL	183	31.51	16.84	27.95	9
c, RC	183	−0.52	−0.27	−0.43	87
c, BLT LASSO	183	−0.42	−0.20	−0.38	77
c, ML-RC	183	−0.98	−0.67	−0.82	77
g, RC	183	−0.84	−0.49	−0.79	90
g, ML-RC	183	−1.30	−0.83	−1.26	80
g, ML-RC, pairwise	183	−1.15	−0.49	−1.00	67
Against the corrected LASSO					
Shrunk target	183	−10.51	−10.03	−7.24	67
Lagged RC	183	283.58	196.85	232.51	1
HAR	183	15.32	11.78	14.76	6
SHAR-F-Exp	30	29.78	12.02	17.38	0
SHAR-F-Exp, corrected	183	7.43	5.65	8.36	9
LASSO	133	2.85	0.00	2.41	1
DCC-NL	183	20.99	8.80	19.82	35
c, RC	183	−11.03	−10.13	−7.66	68
c, BLT LASSO	183	−10.94	−10.32	−7.61	68
c, ML-RC	183	−11.49	−10.57	−8.05	68
g, RC	183	−11.36	−10.45	−8.03	69
g, ML-RC	183	−11.82	−11.09	−8.57	70
g, ML-RC, pairwise	183	−11.66	−11.01	−8.23	70

Monthly difference in the returns likelihood, forecast minus the benchmark named in the panel heading; a negative difference favours the forecast. The third column drops the month with the largest absolute difference, and the last gives the share of months in which the forecast is better.

Table S21 – S&P 500 panel, fifteen-minute grid: how much single months drive the likelihood comparisons

		Returns likelihood difference			
Forecast	Months	Mean	Median	Mean without largest month	Better (% of months)
Against the shrunk target					
Lagged RC	183	709.13	556.38	623.49	0
HAR	183	33.96	30.89	29.96	7
SHAR-F-Exp	67	19.67	22.98	17.97	16
SHAR-F-Exp, corrected	183	18.98	17.92	15.84	18
LASSO	145	25.98	9.62	15.91	34
LASSO, corrected	183	11.74	9.83	7.10	31
DCC-NL	183	31.83	17.84	28.32	8
DCC-NL, monthly	183	17.81	1.66	14.13	30
c, RC	183	−0.51	−0.22	−0.44	87
c, BLT LASSO	183	−0.44	−0.22	−0.42	84
c, ML-RC	183	−0.81	−0.49	−0.70	75
g, RC	183	−0.65	−0.37	−0.63	96
g, ML-RC	183	−1.31	−0.77	−1.27	84
g, ML-RC, pairwise	183	−1.16	−0.82	−1.00	70
Against the corrected LASSO					
Shrunk target	183	−11.74	−9.83	−7.10	69
Lagged RC	183	697.39	541.56	616.39	0
HAR	183	22.22	19.53	21.58	2
SHAR-F-Exp	67	14.27	8.50	12.64	4
SHAR-F-Exp, corrected	183	7.23	6.51	8.74	10
LASSO	145	14.30	0.00	8.33	2
DCC-NL	183	20.08	9.61	18.98	33
DCC-NL, monthly	183	6.07	−0.92	4.77	52
c, RC	183	−12.25	−9.92	−7.54	69
c, BLT LASSO	183	−12.18	−9.94	−7.53	69
c, ML-RC	183	−12.55	−10.26	−7.80	69
g, RC	183	−12.39	−9.96	−7.73	70
g, ML-RC	183	−13.06	−10.28	−8.40	72
g, ML-RC, pairwise	183	−12.90	−10.73	−8.10	71

As Table S20, on the fifteen-minute grid; the month dropped is often March 2020.

Table S22 – S&P 500 panel, five-minute grid: variants of the correlation block

Forecast	Months	Returns lik.	vs corr. LASSO	vs target	Mean α	Indef. (%)	GMV risk
ML-RV volatility, 60-month windows (main)							
c, RC	183	108.42	−11.03 (−2.96)	−0.52 (−3.97)	0.015	0.0	8.23
c, BLT LASSO	183	108.51	−10.94 (−2.97)	−0.42 (−4.27)	0.035	0.0	8.23
c, ML-RC	183	107.96	−11.49 (−3.02)	−0.98 (−4.67)	0.104	0.0	8.17
g, RC	183	108.10	−11.36 (−3.09)	−0.84 (−6.25)	0.025	0.0	8.21
g, ML-RC	183	107.64	−11.82 (−3.27)	−1.30 (−5.55)	0.164	0.0	8.15
g, ML-RC, pairwise	183	107.79	−11.66 (−3.12)	−1.15 (−3.47)	0.064	0.0	8.17
Shrunk target	183	108.94	−10.51 (−2.88)	–	–	0.0	8.29
Drivers from the rescaled measure							
c, RC	183	108.94	−10.51 (−2.88)	0.00 (1.14)	0.000	0.0	8.29
c, BLT LASSO	183	108.49	−10.96 (−2.95)	−0.45 (−3.58)	0.037	0.0	8.23
c, ML-RC	183	108.94	−10.51 (−2.87)	0.01 (1.09)	0.000	0.0	8.29
g, RC	183	108.94	−10.51 (−2.88)	0.00 (4.51)	0.000	0.0	8.29
g, ML-RC	183	108.94	−10.51 (−2.88)	0.00 (4.14)	0.000	0.0	8.29
g, ML-RC, pairwise	183	107.74	−11.71 (−3.04)	−1.20 (−2.86)	0.063	0.0	8.17
Level of the driver kept							
c, RC	183	108.39	−11.06 (−3.09)	−0.55 (−2.16)	0.012	0.0	8.22
c, BLT LASSO	183	108.54	−10.91 (−3.06)	−0.40 (−1.48)	0.027	0.0	8.22
c, ML-RC	183	108.77	−10.68 (−3.07)	−0.17 (−0.25)	0.049	0.0	8.18
g, RC	183	108.39	−11.06 (−3.18)	−0.55 (−1.26)	0.022	0.0	8.19
g, ML-RC	183	108.99	−10.46 (−3.30)	0.05 (0.06)	0.077	0.0	8.16
g, ML-RC, pairwise	183	107.87	−11.58 (−3.99)	−1.07 (−0.82)	0.021	0.0	8.20
RV volatility							
c, RC	183	123.51	−7.84 (−2.09)	−0.62 (−4.64)	0.017	0.0	8.30
c, BLT LASSO	183	123.67	−7.68 (−2.07)	−0.46 (−4.75)	0.038	0.0	8.31
g, RC	183	123.29	−8.06 (−2.17)	−0.84 (−6.64)	0.024	0.0	8.29
g, ML-RC	183	122.89	−8.46 (−2.33)	−1.24 (−5.86)	0.155	0.0	8.24
Shrunk target	183	124.13	−7.22 (−1.95)	–	–	0.0	8.37
36-month windows							
c, RC	183	126.56	−12.22 (−3.20)	−0.00 (−2.14)	0.000	0.0	8.49
c, BLT LASSO	183	126.56	−12.22 (−3.20)	−0.00 (−2.17)	0.000	0.0	8.49
g, RC	183	126.56	−12.22 (−3.20)	−0.00 (−1.46)	0.000	0.0	8.49
g, ML-RC	183	126.42	−12.35 (−3.24)	−0.14 (−2.21)	0.016	0.0	8.49
Shrunk target	183	126.56	−12.22 (−3.20)	–	–	0.0	8.49
Level form							
c, RC	183	108.64	−10.82 (−2.95)	−0.30 (−1.67)	0.002	0.0	8.28
c, BLT LASSO	183	108.95	−10.51 (−2.87)	0.01 (1.08)	0.000	0.0	8.29
c, ML-RC	183	108.62	−10.84 (−2.95)	−0.32 (−1.34)	0.002	0.0	8.29
Sample correlation target							
c, RC	183	208.22	88.76 (6.93)	99.28 (7.28)	0.011	0.0	8.28
c, BLT LASSO	183	211.96	92.51 (6.52)	103.02 (6.75)	0.021	0.0	8.28
c, ML-RC	183	214.16	94.70 (6.77)	105.22 (7.10)	0.052	0.0	8.25

Each block is the variant named in its heading, scored with the volatility path and window of that variant: the returns likelihood per day, the mean difference against BLT's corrected LASSO and against the shrunk target of the same scoring with the Newey–West t statistic in parentheses, the mean weight α on the driver, the share of indefinite months and the annualized risk of the minimum variance portfolio under the rescaled realized covariance. A row starting with c is a link in correlation space and one starting with g a link in coordinates. The main block uses the structure-only drivers of Section 2.3, built from the realized measure before rescaling; in the block with drivers from the rescaled measure they are built after rescaling; in the block with the level of the driver kept, $\mathcal{S}$ is the identity.

Table S23 – S&P 500 panel, fifteen-minute grid: variants of the correlation block

Forecast	Months	Returns lik.	vs corr. LASSO	vs target	Mean α	Indef. (%)	GMV risk
ML-RV volatility, 60-month windows (main)							
c, RC	183	114.82	−12.25 (−2.49)	−0.51 (−4.74)	0.012	0.0	8.46
c, BLT LASSO	183	114.89	−12.18 (−2.51)	−0.44 (−5.61)	0.029	0.0	8.45
c, ML-RC	183	114.52	−12.55 (−2.54)	−0.81 (−5.16)	0.080	0.0	8.40
g, RC	183	114.68	−12.39 (−2.55)	−0.65 (−6.44)	0.015	0.0	8.46
g, ML-RC	183	114.01	−13.06 (−2.69)	−1.31 (−6.21)	0.133	0.0	8.38
g, ML-RC, pairwise	183	114.17	−12.90 (−2.59)	−1.16 (−3.45)	0.068	0.0	8.39
Shrunk target	183	115.33	−11.74 (−2.42)	–	–	0.0	8.51
Drivers from the rescaled measure							
c, RC	183	115.33	−11.74 (−2.42)	0.01 (0.81)	0.000	0.0	8.51
c, BLT LASSO	183	114.84	−12.23 (−2.51)	−0.49 (−5.73)	0.032	0.0	8.45
c, ML-RC	183	115.35	−11.72 (−2.41)	0.02 (1.27)	0.001	0.0	8.51
g, RC	183	115.33	−11.74 (−2.42)	0.00 (4.14)	0.000	0.0	8.51
g, ML-RC	183	115.33	−11.74 (−2.42)	0.00 (3.87)	0.000	0.0	8.51
g, ML-RC, pairwise	183	114.26	−12.81 (−2.59)	−1.07 (−3.26)	0.065	0.0	8.39
Level of the driver kept							
c, RC	183	114.74	−12.33 (−2.56)	−0.58 (−3.03)	0.011	0.0	8.45
c, BLT LASSO	183	114.77	−12.30 (−2.56)	−0.56 (−2.92)	0.028	0.0	8.44
c, ML-RC	183	114.67	−12.40 (−2.63)	−0.66 (−1.20)	0.084	0.0	8.35
g, RC	183	114.68	−12.39 (−2.58)	−0.65 (−2.67)	0.017	0.0	8.44
g, ML-RC	183	114.33	−12.74 (−2.83)	−1.00 (−1.73)	0.121	0.0	8.35
g, ML-RC, pairwise	183	115.21	−11.86 (−2.57)	−0.11 (−0.12)	0.025	0.0	8.46
RV volatility							
c, RC	183	131.87	−9.16 (−1.82)	−0.60 (−5.23)	0.013	0.0	8.59
c, BLT LASSO	183	131.99	−9.04 (−1.82)	−0.48 (−5.66)	0.033	0.0	8.60
g, RC	183	131.81	−9.22 (−1.85)	−0.66 (−6.50)	0.015	0.0	8.61
g, ML-RC	183	131.13	−9.90 (−2.00)	−1.35 (−6.41)	0.132	0.0	8.54
Shrunk target	183	132.48	−8.56 (−1.72)	–	–	0.0	8.66

As Table S22, on the fifteen-minute grid of Bollerslev et al. (2026).

Table S24 – S&P 500 panel, five-minute grid: what a near-singular projection costs

Smallest eigenvalue	Months	Returns likelihood difference in the same months: Projection vs target	Projection vs corrected	Corrected vs target	c, RC vs target
SHAR-F-Exp					
Indefinite	153	–	–	19.2	−0.56
$(0, 0.01)$	1	401.4	348.3	53.1	0.01
$[0.01, 0.05)$	12	40.7	19.3	21.3	−0.31
$[0.05, 0.10)$	7	11.5	5.5	6.0	−0.47
≥ 0.10	10	−1.0	0.0	−1.0	−0.27
LASSO					
Indefinite	50	–	–	20.7	−0.91
$(0, 0.01)$	2	74.1	58.1	16.0	−0.23
$[0.01, 0.05)$	14	20.4	15.3	5.1	−0.36
$[0.05, 0.10)$	13	6.8	3.7	3.0	−0.45
≥ 0.10	104	7.2	0.0	7.2	−0.37

Months grouped by the smallest eigenvalue of the uncorrected projection. In those months, the mean difference in the returns likelihood per day of the uncorrected projection against the shrunk target and against its corrected version, of the corrected projection against the target, and of c, RC against the target; a positive difference favours the second forecast. An indefinite forecast has no likelihood.

Table S25 – S&P 500 panel, five-minute grid: the trigger of the convexity correction

	Correction trigger			
Forecast	0.02	0.05	0.10	0.20
Returns likelihood				
SHAR-F-Exp, corrected	148.18	133.42	126.89	126.24
LASSO, corrected	127.20	122.02	119.45	122.09
Shrunk target	108.94	108.94	108.94	108.94
c, RC	108.42	108.42	108.42	108.42
g, ML-RC	107.64	107.64	107.64	107.64
Difference against the corrected LASSO: returns likelihood (t)				
c, RC	−18.78 (−3.71)	−13.60 (−3.24)	−11.03 (−2.96)	−13.67 (−3.92)
g, ML-RC	−19.56 (−3.95)	−14.39 (−3.52)	−11.82 (−3.27)	−14.46 (−4.23)
Difference against the corrected LASSO: GMV risk, percentage points (t)				
c, RC	−2.81 (−10.70)	−2.14 (−11.78)	−1.93 (−11.49)	−1.41 (−9.16)
g, ML-RC	−2.89 (−11.00)	−2.22 (−12.21)	−2.01 (−11.98)	−1.44 (−9.15)
Months in which the portfolio is corrected (%)				
Shrunk target	0.0	0.0	26.8	94.0
c, RC	0.0	0.0	32.2	94.0
g, ML-RC	0.0	0.0	32.2	95.6

The correction mixes a forecast with the HAR projection of Bollerslev et al. (2026) until its smallest eigenvalue reaches the trigger. It builds the corrected projections and is applied to every forecast before its minimum variance portfolio is formed; 0.1 is the trigger of the other tables, and GMV risk is computed under the rescaled realized covariance. Differences are the forecast minus the corrected LASSO, with Newey–West t statistics in parentheses.

References

Archakov, I. & Hansen, P.R. (2021). A New Parametrization of Correlation Matrices. *Econometrica*, 89, 1699–1715.

Bollerslev, T., Li, S.Z. & Tang, Y. (2026). Forecasting and Managing Correlation Risks. *Management Science*, Articles in Advance, published online 24 April 2026.

Fleming, J., Kirby, C. & Ostdiek, B. (2003). The Economic Value of Volatility Timing Using 'Realized' Volatility. *Journal of Financial Economics*, 67, 473–509.

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.

Newey, W.K. & McFadden, D. (1994). Large Sample Estimation and Hypothesis Testing. in *Handbook of Econometrics*, vol. 4, Elsevier, Amsterdam, 2111–2245.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018). Double/Debiased Machine Learning for Treatment and Structural Parameters. *Econometrics Journal*, 21, C1–C68.

Giacomini, R. & White, H. (2006). Tests of Conditional Predictive Ability. *Econometrica*, 74, 1545–1578.

Pagan, A. (1984). Econometric Issues in the Analysis of Regressions with Generated Regressors. *International Economic Review*, 25, 221–247.