

# *Skeptical review: Factor-Based versus Shrinkage Covariance Estimation for Minimum Variance Portfolios under Heteroskedasticity*

---

## Summary

The paper studies covariance matrix estimation for long-only minimum-variance portfolios (MVPs) in a dynamic, heteroskedastic setting. Using 1,000 daily observations for  $N = 10$  large-cap U.S. equities, returns are filtered via univariate rolling GARCH(1,1) to obtain standardized innovations; covariance matrices are estimated on innovations and then rescaled using one-step-ahead volatility forecasts before solving a constrained MVP each day (Sec. 2.1–2.3). The empirical comparison is between (i) a two-factor covariance model (market factor from PC1 plus an orthogonalized technology long–short sector factor) and (ii) Ledoit–Wolf shrinkage toward a constant-correlation target (Sec. 2.2.1–2.2.2). Reported results suggest shrinkage achieves lower average realized risk and dramatically better numerical conditioning (Sec. 3.1), while the factor model becomes extremely ill-conditioned; the paper argues this is driven more by instability in the idiosyncratic variance component (and its GARCH rescaling) than by lack of factor explanatory power (Sec. 3.2–4).

The question is relevant and the conditioning diagnostics are a valuable lens. However, the current evidence is hard to generalize (single small universe, single window length, one bespoke two-factor specification), several implementation details needed for reproducibility are missing (GARCH, factor construction, Ledoit–Wolf variant, optimizer handling of near-singularity), and there are internal inconsistencies/ambiguities (notably the  $R^2$  values in Fig. 2 vs text; and time indexing around volatility forecasts and rescaling). Strengthening robustness checks, reconciling these inconsistencies, and adding targeted diagnostics to isolate the source of ill-conditioning would substantially improve credibility and interpretability (Sec. 2–4).

## Strengths

- Clear motivation connecting covariance estimation error, heteroskedasticity, and MVP instability, and a well-scoped comparison between a factor structure and shrinkage regularization (Sec. 1).
- A transparent rolling-window pipeline (GARCH filtering  $\rightarrow$  innovation covariance estimation  $\rightarrow$  rescaling  $\rightarrow$  MVP optimization) that mirrors practical workflows for time-varying risk (Sec. 2.1–2.3).
- Useful evaluation dimensions beyond average risk (condition number and turnover), which help connect statistical estimation to numerical stability and trading behavior (Sec. 2.3.2, Sec. 3.1).
- The manuscript attempts a mechanistic explanation (factor span vs idiosyncratic instability) rather than stopping at a pure horse race (Sec. 3.2, Sec. 4).

- Core covariance-model algebra and dimensions are generally consistent (e.g.,  $B_t\Omega_tB_t^\top + \Psi_t$ , and rescaling via  $D\Sigma_zD$ ) and the long-only MVP is correctly posed as a convex QP when  $\Sigma_t$  is PSD (Eqs. (2)–(8), Sec. 2.2–2.3).
- Figures (conceptually) aim to pair performance with diagnostics, which is the right direction for explaining why methods differ (Sec. 3.1–3.2).

## Major issues

1. **External validity is limited by a single very small empirical design ( $N = 10$ , 1,000 days, 60-day rolling window) and one bespoke two-factor specification (PC1 market + a hand-built tech long–short factor) (Sec. 2.1, Sec. 2.2.1, Sec. 3.1–3.2).** With  $N = 10$  the setting is not “high-dimensional,” and factor models typically show their main benefits in larger universes and/or richer factor sets; conversely, the extreme condition numbers reported for the factor model may be idiosyncratic to this universe, window length, and factor definition. As written, conclusions in Sec. 4 can read as broadly ruling out structural factor covariance models under heteroskedasticity, which is stronger than the current evidence supports.

*Recommendation:* Add robustness checks in Sec. 3 that vary at least: (i) window length (e.g., 40/60/120 days), (ii) asset universe size/composition (e.g., expand to 30–50 equities; try a different sector mix or market), and (iii) factor specification complexity (market-only; alternative sector split; optionally a standard style factor if available). Report how realized risk, condition numbers, and turnover move across these variants. If expansion is infeasible, explicitly narrow the claim in Sec. 4 to the studied small-universe/short-window setting and discuss why results might differ for larger universes where factor structure is typically stabilizing.

2. **Key methodological details are missing, limiting reproducibility and making it difficult to assess whether the factor model’s instability is intrinsic or implementation-induced (Sec. 2.1–2.2.2).** Missing/unclear items include: the exact GARCH(1,1) mean specification and innovation distribution (Gaussian vs  $t$ ), parameter constraints, estimation method, and whether GARCH parameters are re-estimated each window (Sec. 2.1); the exact construction of the technology subset and long–short factor (constituents, weights, normalization/standardization, time-invariance) and any subsequent scaling after Gram–Schmidt (Sec. 2.2.1); PCA preprocessing (demeaning, correlation vs covariance matrix); and which Ledoit–Wolf constant-correlation variant/target/intensity formula and software implementation are used (Sec. 2.2.2).

*Recommendation:* Expand Sec. 2.1–2.2 (or add an implementation appendix) specifying: (a) the full GARCH model (mean, distribution, estimation routine, re-estimation frequency, convergence handling); (b) factor definitions with an explicit ticker list for the tech leg, long/short weighting scheme, normalization (e.g., dollar-neutral and unit-variance), and whether factors/loadings are re-scaled after orthogonalization; (c) PCA

computation details (demeaning, matrix choice, sign convention handling); and (d) the exact Ledoit–Wolf reference/variant and how  $\delta_t$  and the constant-correlation target  $F_t$  are computed, including library/code used. This will make the pipeline auditable and help interpret the source of numerical problems.

3. **There is a material inconsistency/ambiguity in the factor-model fit ( $R^2$ ): the text reports  $R^2$  values “mostly between 0.4 and 0.6” (Sec. 3.2), but Fig. 2 (as described in the unstructured report and noted in the structured report) appears to show values around  $\sim 0.83$ – $1.00$ , with spikes to 1.0.** This is central because Sec. 3.2–4 uses  $R^2$  stability/level to argue factor span is adequate and that instability instead comes from  $\Psi_t$  and rescaling. If  $R^2$  is miscomputed, aggregated differently than stated, or affected by leakage/look-ahead, the causal narrative becomes unreliable.

*Recommendation:* Audit and reconcile the  $R^2$  definition and plotting in Sec. 3.2 / Fig. 2: state precisely whether Fig. 2 shows (i) cross-sectional average of per-asset OLS  $R^2$ , (ii) a variance-explained ratio from PCA, or (iii) something else; clarify whether  $R^2$  is in-sample within the 60-day window or evaluated out-of-sample; and report summary stats (mean/median/IQR/min/max) across time and assets. Plot  $R^2$  on the full  $[0, 1]$   $y$ -axis (optionally with an inset) and investigate spikes to exactly 1.0 (potential degenerate windows, near-collinearity, or implementation errors). If any look-ahead is present, correct it and update the conclusions in Sec. 4 accordingly.

4. **Time-indexing around GARCH standardization and rescaling is ambiguous/inconsistent (Sec. 2.1–2.3; Eq. (1) vs narrative; Eqs. (3) and (5)).** The manuscript describes one-step-ahead forecasts (for day  $t + 1$ ) but standardizes as  $z_{i,t} = r_{i,t}/\hat{\sigma}_{i,t}$  (Eq. (1)) and rescales using  $\text{diag}(\hat{\sigma}_t)$  (Eqs. (3)/(5)). Without explicit timing, it is unclear whether  $\Sigma_t$  used for weights targets  $\text{Cov}(r_{t+1}|F_t)$  or  $\text{Cov}(r_t|F_{t-1})$ , and mismatched indices could also contribute to apparent instability.

*Recommendation:* Make timing explicit and consistent throughout Sec. 2: define whether  $\hat{\sigma}_{i,t}$  denotes the conditional s.d. for  $r_{i,t}$  given information at  $t - 1$ , or the forecast for  $r_{i,t+1}$  given information at  $t$ . Then update Eq. (1) and the rescaling in Eqs. (3)/(5) to use matching indices (e.g., use  $\text{diag}(\hat{\sigma}_{t+1|t})$  if  $\Sigma_t$  is meant to forecast next-day return covariance). Add one sentence in Sec. 2.3 clarifying which  $\Sigma$  is optimized to generate  $w_t$  and which realized return  $r_{t+1}$  evaluates it.

5. **The diagnosis of the factor model’s extreme ill-conditioning is plausible but remains largely qualitative and under-identified: it is unclear whether instability originates in (i) innovation-space factor estimation, (ii) near-zero/noisy idiosyncratic variances  $\Psi_t$ , (iii) the GARCH rescaling step amplifying dispersion in vol forecasts, or (iv) PD/regularization/solver handling (Sec. 3.1–3.2, Sec. 4).** The reported mean condition numbers ( $\approx 88,480$  for the factor model) are unusually large for a low-rank-plus-diagonal covariance unless some  $\psi_{i,t}$  are extremely small or numerical handling is problematic.

*Recommendation:* Add targeted diagnostics in Sec. 3.2 to isolate the mechanism: (a) report condition numbers for innovation covariances  $\Sigma_{z,t}$  (before rescaling) for both methods; (b) decompose factor covariance conditioning by reporting  $\kappa(B_t\Omega_tB_t^\top)$ ,  $\kappa(B_t\Omega_tB_t^\top + \Psi_t)$  in innovation space, and  $\kappa$  after rescaling; (c) report the empirical distribution over time of diagonal  $\psi_{i,t}$  (min/percentiles) and of  $\hat{\sigma}_{i,t}$  (min/percentiles), and show whether spikes in  $\kappa$  line up with extreme  $\psi$  or  $\sigma$ ; (d) implement minimal regularizations—e.g., floor  $\psi_{i,t} \geq \epsilon$ , shrink  $\Psi_t$  toward a constant-diagonal target, or smooth  $\Psi_t$  over time—and show the impact on  $\kappa$ , realized risk, and turnover. These additions would convert the narrative in Sec. 4 from conjecture to evidence.

6. **Performance evaluation is not statistically characterized and the realized “variance” metric is potentially misinterpreted (Sec. 2.3.2, Sec. 3.1).** The reported daily realized variance uses  $w^\top r r^\top w = (w^\top r)^2$ , which is a squared realized portfolio return (a second moment), not a variance estimator unless carefully aggregated and mean effects are addressed. In addition, the comparison relies mainly on time-series averages without dispersion measures, confidence intervals, or paired tests, so it is unclear whether the difference (e.g., 0.000126 vs 0.000153) is statistically/economically meaningful.

*Recommendation:* In Sec. 2.3.2, rename the metric as “squared realized return” (or explicitly justify interpreting its time-average as an out-of-sample second moment under a zero-mean approximation). Complement it with a standard out-of-sample variance estimate of portfolio returns over the backtest (or rolling realized variance of portfolio returns). In Sec. 3.1, add dispersion (SD/IQR) for realized risk, condition numbers, and turnover; compute confidence intervals for mean differences (e.g., block bootstrap over days); and run simple paired tests on daily squared returns. Optionally report basic return metrics (mean return, volatility, Sharpe) to contextualize whether lower risk coincides with comparable returns.

7. **Figures and key result presentation contain omissions and potential errors that materially affect interpretability (Sec. 3.1–3.2).** Figure 1 is described as multi-panel (variance/condition number/turnover) but appears incomplete; axis labels/units/time scale are unclear; and condition numbers likely require log scaling to be readable. Figure 2 has the  $R^2$  discrepancy noted above and the  $y$ -axis treatment may visually overstate changes. These presentation issues impede verification of the main claims.

*Recommendation:* Rebuild Figure 1 as a true 3-panel figure (or separate clearly labeled subfigures) with explicit units (daily vs annualized), date axis, and a legend placed outside the plotting area; plot condition numbers on a  $\log_{10}$  scale. For Figure 2, after reconciling  $R^2$ , use the full  $[0,1]$  scale (optionally add an inset), label the  $x$ -axis with dates, and include summary statistics in the caption. Ensure captions state clearly: rolling window length, whether GARCH filtering is applied, and whether quantities are in innovation space or rescaled return space.

8. **Portfolio optimization/PD handling is under-specified despite being central given the paper’s emphasis on ill-conditioning (Sec. 2.3.1, Sec. 3.1).**

With extreme condition numbers, results can depend heavily on whether  $\Sigma_t$  is enforced to be PSD/PD (eigenvalue clipping,  $\epsilon I$  jitter), how the QP is solved, and solver tolerances. Without these details, it is hard to attribute differences to covariance estimators rather than numerical optimization choices.

*Recommendation:* In Sec. 2.3.1, specify the solver/library used for the long-only QP, tolerances, and how non-PD or nearly singular  $\Sigma_t$  is treated (symmetrization, eigenvalue clipping, ridge adjustment  $\epsilon I$ , using singular values for  $\kappa$ ). Report how often PD fixes were needed under each estimator and whether any days were dropped. Consider adding weight-stability diagnostics (max weight, effective number of holdings  $1/\sum w_i^2$ ) to connect ill-conditioning to economically meaningful portfolio concentration beyond turnover.

## Minor issues

1. Dataset description is insufficient for replication and for judging representativeness (Sec. 2.1). The paper does not list tickers, the exact sample period, data source, or return construction (simple vs log; adjusted close/total return), nor handling of missing values/outliers/corporate actions.

*Recommendation:* Augment Sec. 2.1 with a reproducibility table (tickers, sectors, sample start/end dates, data provider), specify return definition and adjustments, and document missing-data/outlier handling. A brief summary of per-asset volatility would also support claims about “high idiosyncratic volatility.”

2. The factor model is described as “structural,” but one factor is statistical (PC1) and the sector factor is bespoke; the economic motivation and relation to standard practice is not fully articulated (Sec. 1, Sec. 2.2.1).

*Recommendation:* Clarify in Sec. 1 and Sec. 2.2.1 that the market factor is PCA-estimated and discuss how this differs from using an observable index. Briefly situate the approach relative to standard models (e.g., Barra-style industry factors, Fama–French) and explain why this particular two-factor design is appropriate for  $N = 10$ .

3. The claim that GARCH-filtered innovations are “approximately homoskedastic” is asserted without supporting diagnostics (Sec. 2.1). Given the centrality of filtering/rescaling, readers need evidence that the step works as intended.

*Recommendation:* Add a short residual diagnostic summary (e.g., ARCH-LM or Ljung–Box on squared standardized residuals per asset; brief plots in an appendix) showing that conditional heteroskedasticity is substantially reduced after filtering.

4. Turnover is discussed but practical frictions (transaction costs) are not incorporated, which limits economic interpretation of turnover differences and “noise-driven trading” claims (Sec. 2.3.2, Sec. 3.1, Sec. 4).

*Recommendation:* Either add a simple proportional transaction cost sensitivity (even a few bps levels) and report net performance impacts, or explicitly state in Sec. 4 that conclusions are conditional on frictionless trading and avoid economically-loaded interpretations of small turnover differences.

5. The condition number definition  $\kappa(\Sigma) = \lambda_{\max}/\lambda_{\min}$  assumes  $\lambda_{\min} > 0$ , but the manuscript does not state what happens when  $\Sigma$  is only PSD or numerically indefinite (Sec. 2.3.2).

*Recommendation:* Add a brief note: either enforce PD before computing  $\kappa$  (and state how), or define  $\kappa$  using singular values to handle PSD/indefinite cases consistently.

6. Notation is occasionally ambiguous between “true” vs “estimated” innovation covariance and method-specific estimators (Sec. 2.2, Eqs. (2)–(5)).

*Recommendation:* Add one sentence (or a small notation table) distinguishing  $\Sigma_{z,t}$  (generic/true concept),  $\hat{\Sigma}_{z,t}^{\text{sample}}$ ,  $\hat{\Sigma}_{z,t}^{\text{factor}}$ , and the corresponding rescaled return covariances used in the MVP,  $\hat{\Sigma}_{z,t}^{\text{LW}}$

## Very minor issues

1. Minor typographical/formatting issues (likely OCR-related): split words (e.g., “environment” in Sec. 4), inconsistent heading formatting (e.g., stray markers around Sec. 2.3.2), and inconsistent capitalization/pluralization of “Minimum Variance Portfolio(s)” across sections.

*Recommendation:* Proofread the final manuscript to fix line breaks, headings, and consistent terminology/capitalization. Prefer display equations with consistent numbering and references (Eq. (1), Eqs. (6)–(8), etc.).

2. Figure captions/title phrasing is sometimes repetitive and terminology occasionally mixes “volatility” and “variance” (Figures around Sec. 3).

*Recommendation:* Tighten captions to state exactly what is plotted (variance vs volatility; innovation vs rescaled return quantities), the rolling window length, and the key takeaway without duplicating main-text narrative.

## Key statements and references

- • The Ledoit–Wolf shrinkage estimator regularizes the sample covariance matrix of innovations,  $S_t$ , by combining it with a constant-correlation target matrix  $F_t$  via the convex combination  $\hat{\Sigma}_{z,t} = (1 - \delta_t)S_t + \delta_t F_t$ , where the

shrinkage intensity  $\delta_t \in [0, 1]$  is computed analytically at each time step to minimize the expected squared error between the estimated and true covariance matrices, rather than being chosen as a free tuning parameter.

- *Reference(s)*: Ledoit, Wolf
- • The Ledoit–Wolf shrinkage estimator, when applied to GARCH(1,1)-filtered innovations within a 60-day rolling window for a universe of ten large-cap U.S. equities, produces Minimum Variance Portfolios with a substantially lower mean out-of-sample realized variance (approximately 0.000126) than a structurally specified two-factor model (approximately 0.000153), indicating superior risk reduction in this empirical setting.
- *Reference(s)*: (none)
- • The covariance matrices produced by the Ledoit–Wolf shrinkage estimator in this study are numerically well-conditioned, with a mean condition number of about 20.8, whereas those produced by the structural two-factor model are severely ill-conditioned, with a mean condition number of about 88,480, implying that the factor-based matrices are nearly singular and that their inverses—and thus the resulting portfolio weights—are highly sensitive to small estimation errors.
- *Reference(s)*: (none)

## Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

**Maths relevance:** light

The paper’s mathematics consists mainly of standard definitions for GARCH-standardized innovations, a two-factor covariance decomposition, Ledoit–Wolf-style shrinkage as a convex combination of sample covariance and a target, a diagonal rescaling to reconstruct return covariances, and a long-only minimum-variance quadratic program. Algebra and matrix dimensions are largely consistent. The main internal problem is inconsistent/ambiguous time indexing for the GARCH volatility forecasts versus the standardization and rescaling equations, which affects whether the covariance used at time  $t$  corresponds to risk on day  $t$  or  $t + 1$ . Several key analytic steps (notably the shrinkage intensity formula) are not shown, limiting verifiability.

### Checked items

1. ✘ **GARCH standardization of returns into innovations** (Eq. (1), Sec. 2.1, p.3)
  - **Claim:** Defines innovations as  $z_{i,t} = r_{i,t}/\hat{\sigma}_{i,t}$  after fitting a rolling GARCH (1,1) and producing a one-step-ahead conditional standard deviation forecast.

- **Checks:** symbol consistency, time-index consistency, dimensional sanity
  - **Verdict:** FAIL; confidence: high; impact: critical
  - **Assumptions/inputs:**  $r_{i,t}$  is a daily return (scalar) for asset  $i$ ,  $\hat{\sigma}_{i,t}$  is a conditional standard deviation associated with  $r_{i,t}$  (must be  $> 0$ ), Standardized innovations are intended to be approximately homoskedastic
  - **Notes:** Text says the procedure generates a one-step-ahead forecast  $\hat{\sigma}_{i,t+1}$ , but Eq. (1) uses  $\hat{\sigma}_{i,t}$  in the denominator. Without an explicit convention re-defining  $\hat{\sigma}_{i,t}$  to mean the forecast for  $t$  (or  $t + 1$ ), the indexing is inconsistent and affects downstream covariance reconstruction and portfolio timing.
2. ✓ **Factor-model covariance decomposition for innovations** (Eq. (2), Sec. 2.2.1, p.3)
- **Claim:** Models innovation covariance as  $\Sigma_{z,t} = B_t \Omega_t B_t^\top + \Psi_t$  with  $\Psi_t$  diagonal.
  - **Checks:** matrix dimension check, notation consistency
  - **Verdict:** PASS; confidence: high; impact: moderate
  - **Assumptions/inputs:**  $B_t$  is  $N \times 2$ ,  $\Omega_t$  is  $2 \times 2$  symmetric,  $\Psi_t$  is  $N \times N$  diagonal with nonnegative entries,  $\Sigma_{z,t}$  is intended to be symmetric positive semidefinite
  - **Notes:** Matrix multiplication is conformable and yields an  $N \times N$  symmetric matrix when  $\Omega_t$  is symmetric. Adding diagonal  $\Psi_t$  preserves symmetry.
3. △ **Orthogonalization implies diagonal factor covariance** (Text in Sec. 2.2.1, p.3)
- **Claim:** Gram–Schmidt orthogonalization of the sector factor w.r.t. the market factor ensures  $\Omega_t$  is diagonal.
  - **Checks:** logical implication check, missing-definition check
  - **Verdict:** UNCERTAIN; confidence: medium; impact: minor
  - **Assumptions/inputs:** Factors are treated as time series within the rolling window,  $\Omega_t$  is computed as the sample covariance of the (possibly demeaned) factor series
  - **Notes:** Gram–Schmidt guarantees orthogonality under a specified inner product, but the paper does not specify whether factors are demeaned and what inner product is used relative to the sample covariance definition. Diagonality of  $\Omega_t$  is verifiable only once these details are specified.
4. △ **Reconstruction of return covariance from innovation covariance (factor model)** (Eq. (3), Sec. 2.2.1, p.4)
- **Claim:** Constructs return covariance as  $\Sigma_{\text{factor},t} = \text{diag}(\hat{\sigma}_t)(B_t \Omega_t B_t^\top + \Psi_t) \text{diag}(\hat{\sigma}_t)$ .
  - **Checks:** algebraic identity check, matrix dimension check, time-index consistency

- **Verdict:** UNCERTAIN; confidence: high; impact: critical
- **Assumptions/inputs:** Return vector  $r_t$  relates to innovation vector  $z_t$  via  $r_t = \text{diag}(\hat{\sigma}_t)z_t$ ,  $\hat{\sigma}_t$  is nonnegative elementwise
- **Notes:** Algebraically correct if  $r_t = Dz_t$  with  $D = \text{diag}(\hat{\sigma}_t)$ . However, combined with Sec. 2.1's 'one-step-ahead' forecast language, it is unclear whether  $\hat{\sigma}_t$  here refers to volatilities for day  $t$  or for day  $t + 1$ . This timing matters for whether  $\Sigma_{\text{factor},t}$  is the covariance used to forecast risk over the next holding period.

5. ✓ **Shrinkage covariance estimator for innovations** (Eq. (4), Sec. 2.2.2, p.4)

- **Claim:** Defines  $\hat{\Sigma}_{z,t} = (1 - \delta_t)S_t + \delta_t F_t$  with  $\delta_t \in [0, 1]$ .
- **Checks:** algebraic form check, matrix dimension check, constraint sanity
- **Verdict:** PASS; confidence: high; impact: moderate
- **Assumptions/inputs:**  $S_t$  and  $F_t$  are  $N \times N$  symmetric matrices,  $\delta_t$  is a scalar shrinkage intensity
- **Notes:** Convex combination is dimensionally consistent. If  $S_t$  and  $F_t$  are PSD and  $\delta_t \in [0, 1]$ , then  $\hat{\Sigma}_{z,t}$  is PSD.

6. △ **Definition of constant-correlation target matrix  $F_t$**  (Text in Sec. 2.2.2, p.4)

- **Claim:**  $F_t$  has sample variances on the diagonal and off-diagonal elements derived from the average pairwise correlation in the window.
- **Checks:** definition completeness, symbol/construct consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: minor
- **Assumptions/inputs:** Average correlation is a scalar  $\bar{\rho}$  computed from the window, Off-diagonal covariances are intended to be  $\bar{\rho}\sqrt{s_{ii}s_{jj}}$  where  $s_{ii}$  are sample variances
- **Notes:** The exact formula for off-diagonal entries is not written. Without it, it is not possible to verify symmetry/PSD properties or confirm the intended mapping from average correlation to covariances.

7. △ **Analytic computation of shrinkage intensity  $\delta_t$**  (Text in Sec. 2.2.2, p.4)

- **Claim:**  $\delta_t$  is calculated analytically at each time step to minimize expected squared error between estimated and true covariance matrices.
- **Checks:** missing-derivation check
- **Verdict:** UNCERTAIN; confidence: high; impact: moderate
- **Assumptions/inputs:** A specific loss function and estimator for the unknown expectations are used
- **Notes:** No analytic expression for  $\delta_t$  (or the minimized objective) is provided in the PDF, so this central step cannot be audited for correctness or for ensuring  $\delta_t \in [0, 1]$  under the implemented formula.

8.  $\triangle$  **Reconstruction of return covariance from innovation covariance (shrinkage model)** (Eq. (5), Sec. 2.2.2, p.4)

- **Claim:** Constructs return covariance as  $\Sigma_{\text{shrinkage},t} = \text{diag}(\hat{\sigma}_t) \hat{\Sigma}_{z,t} \text{diag}(\hat{\sigma}_t)$ .
- **Checks:** algebraic identity check, matrix dimension check, time-index consistency
- **Verdict:** UNCERTAIN; confidence: high; impact: critical
- **Assumptions/inputs:**  $r_t = \text{diag}(\hat{\sigma}_t) z_t$  (consistent with innovation definition)
- **Notes:** Same timing ambiguity as Eq. (3): the rescaling is algebraically correct but the paper does not unambiguously state whether  $\hat{\sigma}_t$  is contemporaneous or one-step-ahead for the holding period  $t \rightarrow t + 1$ .

9.  $\checkmark$  **Minimum variance objective and constraints** (Eqs. (6)–(8), Sec. 2.3.1, p.4)

- **Claim:** Minimize  $w_t^\top \Sigma_t w_t$  subject to  $w_t^\top \mathbf{1} = 1$  and  $w_{i,t} \geq 0$ .
- **Checks:** optimization formulation check, dimension check
- **Verdict:** PASS; confidence: high; impact: moderate
- **Assumptions/inputs:**  $\Sigma_t$  is symmetric positive semidefinite for convexity, Weights are re-optimized each day
- **Notes:** Objective is a scalar quadratic form; constraints are standard for a long-only fully invested MVP.

10.  $\checkmark$  **Realized one-day portfolio variance definition** (Sec. 2.3.2 (Metric 1), p.5)

- **Claim:** Defines realized variance for day  $t + 1$  as  $\sigma_{p,t+1}^2 = w_t^\top r_{t+1} r_{t+1}^\top w_t$ .
- **Checks:** algebraic equivalence check, dimension check
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:**  $r_{t+1}$  is an  $N \times 1$  vector of realized asset returns for day  $t + 1$ ,  $w_t$  is the weight vector chosen at time  $t$
- **Notes:** Expression equals  $(w_t^\top r_{t+1})^2$ , i.e., realized squared portfolio return. Terminology ‘variance’ is acceptable in this one-period realized sense.

11.  $\checkmark$  **Condition number definition for covariance matrices** (Sec. 2.3.2 (Metric 2), p.5)

- **Claim:** Defines  $\kappa(\Sigma) = \lambda_{\max}/\lambda_{\min}$  as the ratio of largest to smallest eigenvalue.
- **Checks:** definition sanity check, edge-case check
- **Verdict:** PASS; confidence: medium; impact: minor
- **Assumptions/inputs:**  $\Sigma$  is symmetric positive definite so that  $\lambda_{\min} > 0$
- **Notes:** Correct for SPD matrices. If  $\Sigma$  is only PSD or becomes indefinite numerically,  $\kappa$  based on eigenvalues may be undefined/negative; the paper does not state handling of this case.

## 12. ✓ Turnover definition (Sec. 2.3.2 (Metric 3), p.5)

- **Claim:** Defines daily turnover as  $\sum_i |w_{i,t} - w_{i,t-1}|$ .
- **Checks:** definition check, constraint compatibility
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:**  $w_t$  and  $w_{t-1}$  are feasible portfolios with same asset universe
- **Notes:** Standard  $L_1$  turnover measure; consistent with long-only weights.

### Limitations

- Only the content present in the provided PDF text/images was used; key analytic steps (e.g., the explicit Ledoit–Wolf shrinkage intensity formula and exact constant-correlation target construction) are not shown and therefore cannot be verified.
- The paper provides definitions but not step-by-step derivations; where claims depend on omitted preprocessing details (e.g., how Gram–Schmidt relates to covariance diagonality), items were marked UNCERTAIN rather than inferred.
- No numerical verification was attempted (per instructions); the audit focuses strictly on symbolic consistency, dimensions, and logical implications.

## Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

All executed numeric checks passed. Verified items were limited to (i) arithmetic relationships among reported summary statistics, (ii) implied rolling-window count from stated panel/window lengths (with convention ambiguity explicitly noted), and (iii) matrix-dimension/logical-constraint consistency checks. Empirical mean quantities stated in the narrative (realized variance, condition number, turnover, and  $R^2$  behavior) could not be recomputed from underlying series based on the provided inputs.

### Checked items

1. ✓ **C1\_mean\_realized\_variance\_ordering\_and\_diff** (Page 5 (Section 3.1, paragraph starting 'Quantitatively...'))
  - **Claim:** Mean realized variance for the shrinkage strategy was approximately 0.000126, whereas the factor-based strategy yielded a higher mean variance of 0.000153.
  - **Checks:** ordering\_and\_difference
  - **Verdict:** PASS
  - **Notes:** Confirmed factor mean variance exceeds shrinkage mean variance; computed absolute difference 0.000027 and relative difference  $\approx 0.2142857$  from the stated values.

2. ✓ **C2\_mean\_condition\_number\_ratio** (Page 6 (Section 3.1, condition number paragraph))
  - **Claim:** Mean condition number is approximately 20.8 (shrinkage) versus approximately 88,480 (factor-based).
  - **Checks:** ratio\_magnitude
  - **Verdict:** PASS
  - **Notes:** Parsed 88,480 as 88480; computed ratio  $88480/20.8 \approx 4253.8462$  and verified it exceeds 1000 as a magnitude sanity check.
3. ✓ **C3\_turnover\_comparison\_and\_diff** (Page 6 (Section 3.1, turnover sentence))
  - **Claim:** Comparable mean turnover (0.197 for the factor model versus 0.214 for the shrinkage model).
  - **Checks:** difference\_and\_relative\_difference
  - **Verdict:** PASS
  - **Notes:** Confirmed shrinkage mean turnover exceeds factor mean turnover; computed absolute difference  $\approx 0.017$  and relative difference  $\approx 0.0862944$  from the stated values.
4. ✓ **C4\_window\_length\_vs\_panel\_length\_feasibility\_count** (Pages 1, 3, 7 (Abstract and Sections 2.1, 4: '1,000-day panel' and '60-day rolling window'))
  - **Claim:** Study uses a 1,000-day panel and fits models on the most recent 60 days in a rolling-window framework.
  - **Checks:** implied\_count\_of\_rolling\_estimations
  - **Verdict:** PASS
  - **Notes:** Computed both common conventions for the number of rolling windows:  $1000 - 60 = 940$  and  $1000 - 60 + 1 = 941$ ; exact indexing convention is not specified in the statement.
5. ✓ **C5\_dimension\_consistency\_factor\_model\_matrices** (Page 3 (Eq. 2 and surrounding definitions))
  - **Claim:**  $B_t$  is  $N \times 2$ ,  $\Omega_t$  is  $2 \times 2$ ,  $\Psi_t$  is  $N \times N$  diagonal;  $\Sigma_{z,t} = B_t \Omega_t B_t^\top + \Psi_t$ .
  - **Checks:** matrix\_dimension\_consistency
  - **Verdict:** PASS
  - **Notes:** With  $N = 10$  and  $K = 2$ , dummy-matrix multiplication yields  $B_t \Omega_t B_t^\top$  as  $10 \times 10$  and confirms conformable addition with  $\Psi_t$  ( $10 \times 10$ ).
6. ✓ **C6\_dimension\_consistency\_rescaling\_equations** (Page 4 (Eq. 3 and Eq. 5))
  - **Claim:**  $\Sigma_{\text{factor},t} = \text{diag}(\hat{\sigma}_t)(\dots)\text{diag}(\hat{\sigma}_t)$  and  $\Sigma_{\text{shrinkage},t} = \text{diag}(\hat{\sigma}_t)\hat{\Sigma}_{z,t}\text{diag}(\hat{\sigma}_t)$ .

- **Checks:** matrix\_dimension\_consistency
- **Verdict:** PASS
- **Notes:** Confirmed  $\text{diag}(\hat{\sigma}_t)$  is  $N \times N$  and  $DAD$  preserves  $N \times N$  shape; also validated the elementwise scaling identity  $\sigma_i \sigma_j A_{ij}$  on a sample entry using dummy data.

7. ✓ **C7\_weight\_constraints\_internal\_consistency** (Page 4 (Eqs. 6–8))

- **Claim:** Long-only MVP constraints:  $w^\top \mathbf{1} = 1$  and  $w_{i,t} \geq 0$  for  $i = 1, \dots, N$ .
- **Checks:** constraint\_implications
- **Verdict:** PASS
- **Notes:** Logical implication check: nonnegative weights summing to 1 imply each weight lies in  $[0, 1]$ . Demonstrated with a constructed example vector and the stated constraints.

### Limitations

- Only parsed text (no numeric tables) is available; most reported empirical metrics (means over the backtest) cannot be recomputed without underlying time series of returns, weights, or covariance estimates.
- Values shown in Figure 1 and Figure 2 are not provided in tabular form; extracting numeric values from plot pixels is out of scope per instructions.
- Several checks are limited to arithmetic, dimensional consistency, and logical implications from stated constraints rather than verifying empirical computations.