
Directionally Correct Views in Portfolio Optimization

Ojas Kumar Sreejit Dey Sarvesh R. Iyer Sandeep Juneja

Ashoka University

{ojas.kumar,sreejit.dey_ug2024,sarvesh.iyer,sandeep.juneja}@ashoka.edu.in

Abstract

Portfolio optimization techniques often use historical data to estimate parameters for Markowitz-type or ML models. Effective allocation requires a way to align the available data with conditions expected over the investment horizon. Inspired by the Black–Litterman framework, we ask whether incorporating expert views leads to better portfolio allocations. In a simple Markowitz setting, we show that even a small amount of directionally correct information can materially improve investment decisions. We also discuss how expert views can be incorporated in end-to-end portfolio models using entropy-based ideas. Somewhat surprisingly, our empirical results indicate that simpler Markowitz-type model perform as well as an end-to-end model, once directionally correct experts are incorporated.

1 Introduction

Portfolio design is a historically rich field, combining simple financial insights with complex mathematical models. A seminal contribution to this field was made by Markowitz [1], who first introduced a quantitative framework in financial investment eventually leading to the Capital Asset Pricing Model (CAPM) [2, 3, 4]. A major drawback of the theory lay in the inability to predict the mean μ and covariance Σ of future asset returns, and the extreme sensitivity of the optimal allocation to any estimation errors [5, 6, 7, 8, 9, 10, 11].

Black and Litterman [12] in their pioneering work developed a Bayesian framework to arrive at forward looking mean–variance estimates by incorporating qualitative *expert views* into a quantitative framework. This work inspired many subsequent works in portfolio optimization as well as in option pricing, e.g., [13, 14, 15, 16, 17, 18, 19, 20]. Precisely what quantitative factors determine the effectiveness of expert views is a question of importance. More recently, data-driven approaches have sought to avoid estimating μ and Σ [21, 22], raising the question of how forward-looking expert views can be incorporated directly into an end-to-end model.

Overview. In this article, we provide quantitative estimates showing that even a modest amount of *directionally correct* information in expert views can substantially improve portfolio allocations, and we develop a framework for incorporating such views into both Markowitz and end-to-end models. § 2 shows that the Sharpe ratio increases monotonically, and is concave near the optimum, as the directionally correct view strengthens, and § 3 shows this holds robustly under small errors in the view. § 4 extends this notion to an end-to-end, ML-based portfolio model and finds empirically that it performs comparably to Markowitz once directionally correct views are incorporated. We use the Sharpe ratio throughout as our measure of portfolio performance [23, 24].

2 Sharpe Ratio under Directionally Correct Views

Under the standard Markowitz framework, we consider n assets with random return vector $\mathbf{r} = (r_1, \dots, r_n)^\top$. Its mean and covariance are denoted by $\boldsymbol{\mu} := \mathbb{E}[\mathbf{r}]$, and $\boldsymbol{\Sigma} := \text{Cov}(\mathbf{r})$. A portfolio $\mathbf{w} \in \mathbb{R}^n$ has return $r_{\mathbf{w}} = \mathbf{w}^\top \mathbf{r}$, with $\mathbb{E}[r_{\mathbf{w}}] = \mathbf{w}^\top \boldsymbol{\mu}$, and $\text{Var}(r_{\mathbf{w}}) = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$. We measure the performance of a portfolio by the widely-used Sharpe ratio,

$$\text{SR}(\mathbf{w}) = \frac{\mathbf{w}^\top \boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}}. \quad (1)$$

Here, r_f is the risk-free rate, which we will assume to be 0 without loss of generality. We refer to the portfolio maximizing (1) as the *maximum-Sharpe portfolio*. When short-selling is allowed and $\boldsymbol{\Sigma}$ is invertible, the maximum-Sharpe portfolio has weights $\mathbf{w}^* \propto \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, where $\mathbf{1}$ denotes the vector of ones; see [25].

Suppose an expert's prediction of the future mean and covariance are $\boldsymbol{\nu}$ and $\boldsymbol{\Omega} \succ 0$ respectively. To determine how this portfolio performs under the distribution that actually governs returns, we calculate the realized Sharpe ratio and maximum possible Sharpe ratio as follows:

$$S_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\boldsymbol{\nu}, \boldsymbol{\Omega}) := \frac{\boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\nu}}{\sqrt{\boldsymbol{\nu}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Omega}^{-1} \boldsymbol{\nu}}}, \quad S_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^* := S_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sqrt{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}. \quad (2)$$

Proposition 1 (Sharpe loss as an angle). *Assume $\boldsymbol{\mu}, \boldsymbol{\nu} \neq 0$ and $\boldsymbol{\Sigma}, \boldsymbol{\Omega} \succ 0$. There exists a unique angle $\theta_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\boldsymbol{\nu}, \boldsymbol{\Omega}) \in [0, \pi]$ such that $\cos \theta_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\boldsymbol{\nu}, \boldsymbol{\Omega}) = \frac{\boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\nu}}{\sqrt{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \boldsymbol{\nu}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Omega}^{-1} \boldsymbol{\nu}}}$. Moreover, we have $S_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\boldsymbol{\nu}, \boldsymbol{\Omega}) = S_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^* \cos \theta_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\boldsymbol{\nu}, \boldsymbol{\Omega})$, so the Sharpe ratio is maximized when $\boldsymbol{\Omega}^{-1} \boldsymbol{\nu} = c \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ for some $c > 0$.*

We say that an expert view is *directionally correct* if, when $\boldsymbol{\mu}$ is the true return mean and $\boldsymbol{\nu}$ is our present forecast, the expert's prediction is of the form $\boldsymbol{\nu}_\lambda := (1 - \lambda)\boldsymbol{\nu} + \lambda\boldsymbol{\mu}$, for some $\lambda \in [0, 1]$. For the next result, we assume that our original prediction, the expert view, as well as the real covariance matrix are the same (namely $\boldsymbol{\Sigma}$). We define $A := \langle \boldsymbol{\nu}, \boldsymbol{\nu} \rangle_{\boldsymbol{\Sigma}^{-1}}$, and $B := \langle \boldsymbol{\nu}, \boldsymbol{\mu} - \boldsymbol{\nu} \rangle_{\boldsymbol{\Sigma}^{-1}}$, where $\langle \mathbf{x}, \mathbf{y} \rangle_{\boldsymbol{\Sigma}^{-1}} := \mathbf{x}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}$.

By Theorem 2, $S(\lambda)$ is concave on the entire interval $[0, 1]$ if and only if $A + 3B \geq 0$; since $A \geq 0$, this holds whenever $B = \langle \boldsymbol{\nu}, \boldsymbol{\mu} - \boldsymbol{\nu} \rangle_{\boldsymbol{\Sigma}^{-1}} \geq 0$, equivalently $\cos \theta' = \frac{\langle \boldsymbol{\nu}, \boldsymbol{\mu} - \boldsymbol{\nu} \rangle_{\boldsymbol{\Sigma}^{-1}}}{\|\boldsymbol{\nu}\|_{\boldsymbol{\Sigma}^{-1}} \|\boldsymbol{\mu} - \boldsymbol{\nu}\|_{\boldsymbol{\Sigma}^{-1}}} > 0$. Financially, this assumption means the future view remains sufficiently close to the historical mean. Figure 1b plots the realized Sharpe ratio against λ on Dataset 1 (§ B.2), for which this condition holds; the observed concavity around the maximum substantiates our theoretical claim. When $A + 3B < 0$, concavity need not hold over the full interval, only around $\lambda = 1$ as Theorem 2 guarantees, and Figure 1a shows the resulting inflection point where the curve transitions from convex to concave.

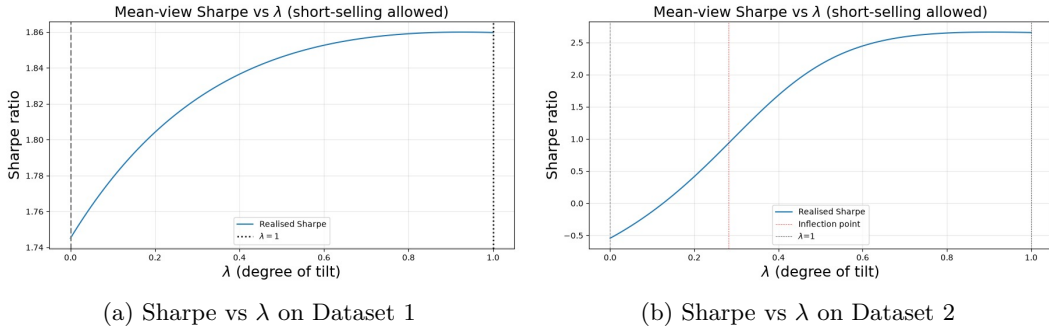


Figure 1: Concavity of Sharpe as a function of λ

We prove Proposition 1 and an analogous statement for covariance views as well as Theorem 2 in § A.

3 Sensitivity of View Incorporation in the Markowitz Framework

In this section, we explain how expert views on the mean of returns can be theoretically incorporated in the Markowitz framework. We also state the perturbation bound Theorem 2, subsequently proved in § A.2. Throughout this section, we assume the existence of a historical measure P on returns (see § B.1).

In the mean-view setup, an expert specifies the mean $\boldsymbol{\mu}^{\text{view}}$ of the future distribution. We incorporate the expert view as in [12]. For two probability distributions Q and P with densities q and p on the same support, the Kullback-Leibler (KL) divergence of Q from P is defined as

$$D_{KL}(Q \| P) = \mathbb{E}_Q \left[\log \frac{q(\mathbf{r})}{p(\mathbf{r})} \right] = \int q(\mathbf{r}) \log \frac{q(\mathbf{r})}{p(\mathbf{r})} d\mathbf{r}.$$

Incorporation is performed by changing the historical measure P to Q , where Q is the solution to the constrained problem $Q^* = \operatorname{argmin}_{Q: \mathbb{E}_Q[\mathbf{r}] = \boldsymbol{\mu}^{\text{view}}} D_{KL}(Q \| P)$. If we assume that $\mathbf{r}_t \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ under P , then it is well-known that $\mathbf{r}_t \sim \mathcal{N}(\boldsymbol{\mu}^{\text{view}}, \boldsymbol{\Sigma})$ under Q . Theorem 2 applies to this particular scenario. However, under non-Gaussian returns, a mean view may also perturb the covariance matrix. We show that this perturbation costs only a *second-order* loss in Sharpe ratio, assuming that the perturbation is small enough relative to the sizes of the true mean and variance. In this sense, our earlier results on directionally correct mean views are robust.

Recall $S_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$ and $S_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^*$ from (2). For a symmetric matrix A , let $\|A\|_{op} = \sup_{\mathbf{x} \in \mathbb{R}^n \setminus \{0\}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$.

Theorem 2. *Let $\boldsymbol{\Sigma}, \boldsymbol{\Omega} \succ 0$, and $\boldsymbol{\mu}, \boldsymbol{\nu}$ be fixed. For some $m > 0$, suppose that $\|\boldsymbol{\mu}\| \geq m$, $\boldsymbol{\Sigma} - mI \succ 0$ hold. Furthermore, assume that $\|\boldsymbol{\mu} - \boldsymbol{\nu}\| < \frac{m}{2}$, $\|\boldsymbol{\Omega} - \boldsymbol{\Sigma}\|_{op} < \frac{m}{2}$. Then,*

$$0 \leq \frac{S_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^* - S_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\boldsymbol{\nu}, \boldsymbol{\Omega})}{S_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^*} < \frac{2\|\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{op}^4}{m^2} (\|\boldsymbol{\mu} - \boldsymbol{\nu}\| + \|\boldsymbol{\Omega} - \boldsymbol{\Sigma}\|_{op})^2.$$

4 Expert Views in the End-to-End Framework

The end-to-end approach to portfolio optimization infers asset allocations directly from historical data, bypassing the intermediate step of estimating the return distribution [21, 22]. We incorporate mean views into this framework through an exponential change of measure. A neural network f_ϕ maps past market information to portfolio weights: the weight selected on day t is a function of the L -day lookback window preceding the decision date.

We work with n assets and $T+L$ days of return data $x_t = (x_{t,1}, \dots, x_{t,n})^\top$, $t = -L+1, \dots, T$. For each day $i = 0, \dots, T-1$, the network outputs the next day's portfolio weight $w_{i+1} = f_\phi(x_{i-L+1}, \dots, x_i)$, yielding return $R_i(\phi) = x_{i+1}^\top w_{i+1}$. Following [21], ϕ is trained by gradient ascent directly on the Sharpe ratio of these portfolio returns, rather than an intermediate return-prediction loss. The only difference being that we modify the Sharpe ratio via likelihood ratios to incorporate directionally correct views.

Let $\boldsymbol{\mu}^{\text{hist}} = \mathbb{E}_P[\mathbf{r}]$ be the mean under the empirical distribution P (see § B.1 for its construction) and $\boldsymbol{\mu}^{\text{view}}$ the mean implied by the expert view; for $\lambda \in [0, 1]$, set $\boldsymbol{\mu}_\lambda^{\text{view}} = (1-\lambda)\boldsymbol{\mu}^{\text{hist}} + \lambda\boldsymbol{\mu}^{\text{view}}$. As in § 3, we reweight via the likelihood ratio

$$\ell_\lambda(\mathbf{r}) = \frac{\exp(\boldsymbol{\eta}_\lambda^\top \mathbf{r})}{\mathbb{E}_P[\exp(\boldsymbol{\eta}_\lambda^\top \mathbf{r})]}, \quad \mathbb{E}_P[\ell_\lambda(\mathbf{r})\mathbf{r}] = \boldsymbol{\mu}_\lambda^{\text{view}}.$$

Since P is uniform on $\mathbf{r}_1, \dots, \mathbf{r}_T$, this gives the explicit weights $q_{t,\lambda} = \exp(\boldsymbol{\eta}_\lambda^\top \mathbf{r}_t) / \sum_{s=1}^T \exp(\boldsymbol{\eta}_\lambda^\top \mathbf{r}_s)$, with $\sum_{t=1}^T q_{t,\lambda} \mathbf{r}_t = \boldsymbol{\mu}_\lambda^{\text{view}}$.

With $m_\lambda(\phi) = \mathbb{E}_P[\ell_\lambda(\mathbf{r}_t)R_t(\phi)]$ and $v_\lambda(\phi) = \mathbb{E}_P[\ell_\lambda(\mathbf{r}_t)(R_t(\phi) - m_\lambda(\phi))^2]$, we update ϕ by gradient ascent on the view-adjusted Sharpe ratio $S_{\phi,\lambda} = m_\lambda(\phi) / \sqrt{v_\lambda(\phi)}$.

Experiments. We test this framework against the Markowitz model of § 3 on daily NSE India data for the 20 largest firms by market capitalization (2006–2025; see § B.2). Both models are given a view $\boldsymbol{\mu}_\lambda$, $\lambda \in \{0, 0.1, \dots, 1\}$, interpolating between the historical mean

(2006–2023) and the realized mean over an evaluation period (2024–2025), and are evaluated on that period; $\lambda = 1$ represents a perfectly accurate view unavailable at construction time. Given constraints on short-selling in Indian markets, we restrict both experiments to long-only portfolios, giving a common constraint under which to compare the two approaches.

Models. For Markowitz, the tilted probabilities $q_t(\lambda)$ that impose μ_λ also determine a covariance estimate¹ $\Sigma_\lambda = \sum_{t=1}^T q_t(\lambda)(\mathbf{r}_t - \mu_\lambda)(\mathbf{r}_t - \mu_\lambda)^\top$; using $(\mu_\lambda, \Sigma_\lambda)$, we solve for the long-only maximum-Sharpe weights $\mathbf{w}_\lambda \in \arg \max_{\mathbf{w} \geq 0, \mathbf{1}^\top \mathbf{w} = 1} \mathbf{w}^\top \mu_\lambda / \sqrt{\mathbf{w}^\top \Sigma_\lambda \mathbf{w}}$. For the end-to-end model, the network takes a lookback window of past returns and outputs a long-only, fully invested portfolio through a final softmax layer; the reported architecture is a single-layer LSTM with 16 hidden units, chosen after exploratory experiments with fully connected networks, CNNs, and larger or deeper LSTMs, none of which improved on it, consistent with the findings of [21]. For each λ , we train 100 random initializations under an identical protocol: 300 full-batch updates, with a 50-update warm-up followed by cosine annealing over the remaining 250.

Results. Figure 2 compares the Markowitz portfolio with the mean performance of the 100 LSTM runs, reported after a one-way transaction cost of 20 bps on the value traded.²³ Both improve sharply as λ increases, with much of the gain arriving already at small λ and tapering off as the view approaches the future mean; the two Sharpe curves are strikingly similar throughout, and dispersion across LSTM seeds falls sharply for more informative views. After costs, the LSTM performs slightly worse than Markowitz, particularly at intermediate λ , consistent with its greater turnover. Exact Sharpe ratios and drawdowns for each λ , along with portfolio allocations at $\lambda = 0, 0.5, 1$, are given in § B.3.

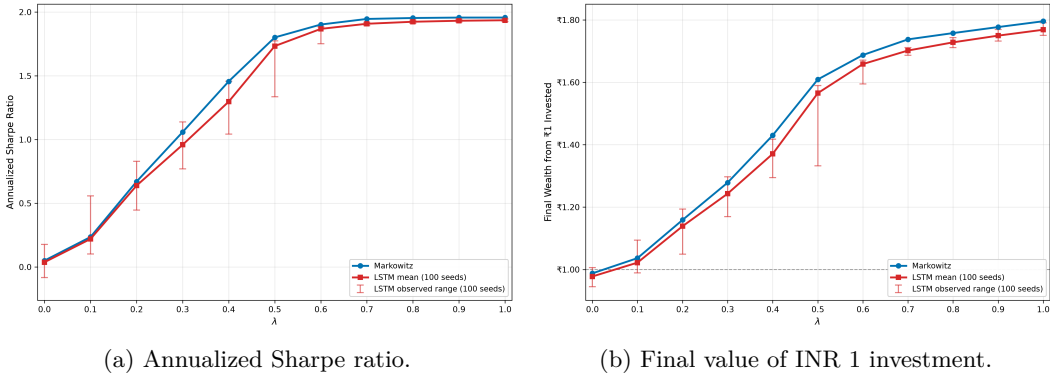


Figure 2: Performance of Markowitz vs End-To-End across λ .

5 Conclusion

In this article, we demonstrated that incorporating directionally correct expert views can improve portfolio design. We also showed that the Sharpe ratio is stable to minor errors in these views, heuristically demonstrating why complex ML models learn allocations similar to those predicted by Markowitz theory and, in turn, achieve comparable Sharpe ratios in our experiments.

Having established the value of directionally correct views, a natural next step is producing such views in practice, since our experiments constructed them using future information unavailable at decision time. We leave this to future work.

¹We estimate Σ_λ via the tilted probabilities $q_t(\lambda)$ rather than by bootstrapping. We found this to be less noisy in practice.

²This is a conservative estimate: delivery-based equity transactions in India carry a securities transaction tax of 10 bps on both purchases and sales, before brokerage and other costs.

³If $\mathbf{w}_t$ is the target allocation and $\mathbf{w}_{t-1}$ the allocation immediately before rebalancing, the cost deducted from the daily return is $c \sum_{i=1}^n |w_{t,i} - w_{t-1,i}|$.

References

- [1] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.
- [2] William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.
- [3] John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37, 1965.
- [4] Jan Mossin. Equilibrium in a capital asset market. *Econometrica*, 34(4):768–783, 1966.
- [5] Michael J. Best and Robert R. Grauer. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies*, 4(2):315–342, 1991.
- [6] Vijay K. Chopra and William T. Ziemba. The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, 19(2):6–11, 1993.
- [7] Philippe Jorion. Bayes–stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3):279–292, 1986.
- [8] Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.
- [9] Yves Choueifaty and Yves Coignard. Toward maximum diversification. *The Journal of Portfolio Management*, 35(1):40–51, 2008.
- [10] Sébastien Maillard, Thierry Roncalli, and Jérôme Teïletche. The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management*, 36(4):60–70, 2010.
- [11] Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683, 2003.
- [12] Fischer Black and Robert Litterman. Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43, 1992.
- [13] Peter Buchen and Michael Kelly. The maximum entropy distribution of an asset inferred from option prices. *The Journal of Financial and Quantitative Analysis*, 31(1):143–159, March 1996.
- [14] Marco Avellaneda, Robert Buff, Craig Friedman, Nicolas Grandchamp, Łukasz Kruk, and Joshua Newman. Weighted monte carlo: A new technique for calibrating asset pricing models. *International Journal of Theoretical and Applied Finance*, 4(1):91–119, March 2001.
- [15] Michael Stutzer. A simple nonparametric approach to derivative security valuation. *The Journal of Finance*, 51(5):1633–1652, December 1996.
- [16] Attilio Meucci. Beyond Black-Litterman: Views on Non-Normal Markets. *SSRN Electronic Journal*, 2005. Working Paper.
- [17] Attilio Meucci. Fully flexible views: Theory and practice. *Risk*, 21(10):97–102, 2008.
- [18] Jacques Pézier. Global portfolio optimization revisited: A least discrimination alternative to black-litterman. ICMA Centre Discussion Papers in Finance icma-dp2007-07, ICMA Centre, University of Reading, July 2007.
- [19] Edward Qian and Stephen A. Gorman. Conditional distribution in portfolio theory. *Financial Analysts Journal*, 57(2):44–51, 2001.

- [20] Joseph Simonian. Copula-opinion pooling with complex opinions. *Quantitative Finance*, 14(6):941–946, 2014.
- [21] Zihao Zhang, Stefan Zohren, and Stephen Roberts. Deep learning for portfolio optimization. *The Journal of Financial Data Science*, 2(4):8–20, 2020.
- [22] Chao Zhang, Zihao Zhang, Mihai Cucuringu, and Stefan Zohren. A universal end-to-end approach to portfolio optimization via deep learning. *arXiv preprint arXiv:2111.09170*, 2021.
- [23] William F. Sharpe. Mutual fund performance. *The Journal of Business*, 39(1):119–138, 1966.
- [24] William F. Sharpe. The sharpe ratio. *The Journal of Portfolio Management*, 21(1):49–58, 1994.
- [25] Robert C. Merton. An analytic derivation of the efficient portfolio frontier. *The Journal of Financial and Quantitative Analysis*, 7(4):1851–1872, 1972.
- [26] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning series. MIT Press, Cambridge, MA, 2012.

A Theoretical results and proofs

In this section, we state and prove the theoretical results in this article, including any that were mentioned in the main body.

A.1 Proofs in Section 2

In this section, we prove the results in Section 2. Recall the definition of the Sharpe ratio (1), and that (μ, Σ) are the true mean and variance of future returns, while (ν, Ω) are the mean and variance estimated by any method.

A.1.1 Proof of Proposition 1 and Theorem 2

We begin with the proof of Proposition 1

Proof of Proposition 1. Since $\Sigma \succ 0$, the bilinear form $\langle \mathbf{x}, \mathbf{y} \rangle_{\Sigma} := \mathbf{x}^{\top} \Sigma \mathbf{y}$ defines an inner product on $\mathbb{R}^n$. By the Cauchy-Schwarz inequality applied to this inner product, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$|\mathbf{x}^{\top} \mathbf{y}| \leq \sqrt{\mathbf{x}^{\top} \Sigma^{-1} \mathbf{x}} \sqrt{\mathbf{y}^{\top} \Sigma \mathbf{y}}.$$

Taking $\mathbf{x} = \mu$ and $\mathbf{y} = \Omega^{-1} \nu$ we have

$$\left| \frac{\mu^{\top} \Omega^{-1} \nu}{\sqrt{\mu^{\top} \Sigma^{-1} \mu} \sqrt{\nu^{\top} \Omega^{-1} \Sigma \Omega^{-1} \nu}} \right| \leq 1.$$

Since $\mu, \nu \neq 0$ and $\Sigma, \Omega \succ 0$, both terms in the denominator are strictly positive. Thus there is a unique $\theta_{\mu, \Sigma}(\nu, \Omega) \in [0, \pi]$ whose cosine is the quantity above. Equivalently, this is the angle under $\langle \cdot, \cdot \rangle_{\Sigma}$ between $\Sigma^{-1} \mu$ and $\Omega^{-1} \nu$.

Now, through minor calculations, it is easily verified that $S_{\mu, \Sigma}(\nu, \Omega) = S_{\mu, \Sigma}^* \cos \theta_{\mu, \Sigma}(\nu, \Omega)$. Therefore $S_{\mu, \Sigma}(\nu, \Omega) \leq S_{\mu, \Sigma}^*$, with equality if and only if the angle is zero. By the equality condition in Cauchy-Schwarz, this occurs iff $\Omega^{-1} \nu = c \Sigma^{-1} \mu$ for some $c > 0$. $\square$

Next, we prove Theorem 2.

Proof of Theorem 2. Using $\nu_{\lambda} = \nu + \lambda(\mu - \nu)$, we can write

$$S(\lambda) = \frac{A + B + \lambda(B + C)}{(A + 2\lambda B + \lambda^2 C)^{1/2}} \implies S'(\lambda) = \frac{(1 - \lambda)(AC - B^2)}{(A + 2\lambda B + \lambda^2 C)^{3/2}}.$$

By the Cauchy-Schwarz inequality, $B^2 \leq AC$ with equality if and only if ν and $\mu - \nu$ are collinear, which is equivalent to μ and ν being collinear. Hence $S'(\lambda) \geq 0$ on $[0, 1]$ and $S'(1) = 0$.

For concavity, differentiating once more yields

$$S''(\lambda) = - \frac{(AC - B^2)H(\lambda)}{(A + 2\lambda B + \lambda^2 C)^{5/2}}, \quad (3)$$

where $H(\lambda) = A + 3B + (3C - B)\lambda - 2C\lambda^2$. Since $C \geq 0$, H is concave in λ . At the endpoints, $H(0) = A + 3B$, while $H(1) = \langle \mu, \mu \rangle_{\Sigma^{-1}} > 0$. Therefore, $H(\lambda) \geq 0$ for every $\lambda \in [0, 1]$ if and only if $A + 3B \geq 0$.

In the non-collinear case, $AC - B^2 > 0$, hence $S''(\lambda) < 0$ throughout $[0, 1]$ by (3). $\square$

A.1.2 Covariance Views

In this section, we describe an analogue of Theorem 2 for covariance views. First, we describe how to integrate these views in the Gaussian setting. Then, we state our main result, Theorem 4 and prove it.

For a fixed symmetric matrix A and $k \in \mathbb{R}$, consider a view of the form $\text{tr}(A\Omega) = k$, where Ω is the covariance matrix of future returns. Interestingly, for Gaussian returns, incorporating this view corresponds to a linear shift of the inverse matrix.

Proposition 3. *Let $\Sigma \succ 0$ and consider*

$$\Omega^* = \arg \min_{\Omega \succ 0} D_{\text{KL}}(\mathcal{N}(\mu, \Omega) \parallel \mathcal{N}(\mu, \Sigma)) \quad \text{subject to} \quad \text{tr}(A\Omega) = k,$$

for some fixed symmetric A and $k \in \mathbb{R}$. Then $\Omega^* = (\Sigma^{-1} + \lambda A)^{-1}$, for some scalar $\lambda \in \mathbb{R}$ determined by the constraint.

Proof. An explicit formula for the KL divergence between multivariate Gaussians can be found in [26, (21.171)]. Using this formula,

$$D_{\text{KL}}(\mathcal{N}(\mu, \Omega) \parallel \mathcal{N}(\mu, \Sigma)) = \frac{1}{2} \left(\text{tr}(\Sigma^{-1}\Omega) - \log \det(\Sigma^{-1}\Omega) - n \right).$$

Thus, it is sufficient to minimize $f(\Omega) = \text{tr}(\Sigma^{-1}\Omega) - \log \det \Omega$, which is strictly convex on $\{\Omega \succ 0\}$. We form the Lagrangian

$$\mathcal{L}(\Omega, \lambda) = \text{tr}(\Sigma^{-1}\Omega) - \log \det \Omega + \lambda (\text{tr}(A\Omega) - k).$$

Using the matrix identities $\partial_{\Omega} \text{tr}(\Sigma^{-1}\Omega) = \Sigma^{-1}$ and $\partial_{\Omega} \log \det \Omega = \Omega^{-1}$, the stationarity condition $\partial_{\Omega} \mathcal{L} = 0$ gives

$$\Sigma^{-1} - \Omega^{-1} + \lambda A = 0 \quad \Longleftrightarrow \quad \Omega^{-1} = \Sigma^{-1} + \lambda A.$$

Since f is strictly convex, this stationary point is the unique minimizer whenever the constraint is feasible, completing the proof. $\square$

Now, let $P = \mathcal{N}(\mu, \Sigma)$ be our estimate and $Q = \mathcal{N}(\mu, \Omega)$ be the true future distribution of asset returns. As for mean views, we call a covariance view as directionally correct if the forecast is of the form $\Omega(t) = (\Sigma^{-1} + t\Delta)^{-1}$ for some $t \in [0, 1]$, where $\Delta := \Omega^{-1} - \Sigma^{-1}$.

By Proposition 1 the optimal allocation has weights $\mathbf{w}(t) \propto \Omega(t)^{-1}\mu$. Let $S(t) = S_{\mu, \Sigma}(\mu, \Omega(t))$. The following analogue of Theorem 2 exists in this setting.

Theorem 4 (Monotonicity and concavity). *Let the Sharpe realized under the true distribution $\mathcal{N}(\mu, \Omega)$, of the tangency portfolio $\mathbf{w}(t) \propto \Omega(t)^{-1}\mu$, be denoted by*

$$S(t) := S_{\mu, \Omega}(\mu, \Omega(t))$$

Then $S(t)$ is nondecreasing on $[0, 1]$. Further, if $\mathbf{p}$ and $\mathbf{q}$ are not collinear, then $S(t)$ is strictly increasing on $[0, 1]$. Moreover, $S(t)$ is concave on $[0, 1]$ if and only if $A + 3B \geq 0$; in particular $B \geq 0$ is sufficient for concavity.

Prior to proving the theorem, we introduce the notation $A := \langle \mathbf{p}, \mathbf{p} \rangle_{\Omega}$, $B := \langle \mathbf{p}, \mathbf{q} - \mathbf{p} \rangle_{\Omega}$, $C := \langle \mathbf{q} - \mathbf{p}, \mathbf{q} - \mathbf{p} \rangle_{\Omega}$, where $\mathbf{p} := \Sigma^{-1}\mu$, $\mathbf{q} := \Omega^{-1}\mu$, $\langle \mathbf{x}, \mathbf{y} \rangle_{\Omega} := \mathbf{x}^{\top} \Omega \mathbf{y}$.

Proof. Since $\Delta\mu = \Omega^{-1}\mu - \Sigma^{-1}\mu = \mathbf{q} - \mathbf{p}$, the defining relation for $\Omega(t)^{-1}$ gives

$$\Omega(t)^{-1}\mu = \Sigma^{-1}\mu + t\Delta\mu = \mathbf{p} + t(\mathbf{q} - \mathbf{p}) =: \mathbf{p}_t,$$

This is exactly the structure of ν_{λ} in Theorem 2. Assuming $\Omega(t) \succ 0$ and $\mu \neq 0$, we have $\mathbf{p}_t = \Omega(t)^{-1}\mu \neq 0$ for every $t \in [0, 1]$. By (2),

$$S(t) = \frac{\mu^{\top} \mathbf{p}_t}{\sqrt{\mathbf{p}_t^{\top} \Omega \mathbf{p}_t}}.$$

Using $\Omega \mathbf{q} = \mu$, the numerator satisfies $\mu^{\top} \mathbf{p}_t = \mathbf{q}^{\top} \Omega \mathbf{p}_t = \langle \mathbf{q}, \mathbf{p}_t \rangle_{\Omega}$,

$$S(t) = \frac{\langle \mathbf{q}, \mathbf{p}_t \rangle_{\Omega}}{\sqrt{\langle \mathbf{p}_t, \mathbf{p}_t \rangle_{\Omega}}} = \frac{A + B + t(B + C)}{(A + 2tB + t^2C)^{1/2}},$$

where the second equality follows from expanding $\mathbf{p}_t = \mathbf{p} + t(\mathbf{q} - \mathbf{p})$ in the inner product $\langle \cdot, \cdot \rangle_{\Omega}$ and using the definitions of A, B, C . Differentiating,

$$S'(t) = \frac{(1-t)(AC - B^2)}{(A + 2tB + t^2C)^{3/2}}.$$

This is exactly the expression for $S'(\lambda)$ obtained in the proof of Theorem 2. Thus, the same analysis follows, giving us monotonicity of the Sharpe ratio as a function of t , and concavity iff $A + 3B \geq 0$. $\square$

A.2 Proof of Theorem 2

Proof of Theorem 2. Suppose that $\mu, \nu, \Sigma, \Omega, m$ are as in the statement of the theorem. By Proposition 1 we have

$$\frac{S_{\mu, \Sigma}^* - S_{\mu, \Sigma}(\nu, \Omega)}{S_{\mu, \Sigma}^*} = 1 - \cos \theta_{\mu, \Sigma}(\nu, \Omega), \quad (4)$$

where

$$\cos \theta_{\mu, \Sigma}(\nu, \Omega) = \frac{\mu^\top \Omega^{-1} \nu}{\sqrt{\mu^\top \Sigma^{-1} \mu} \sqrt{\nu^\top \Omega^{-1} \Sigma \Omega^{-1} \nu}}. \quad (5)$$

The left side of the inequality in the theorem follows directly from Proposition 1. Thus, we must only prove the right hand side.

To do this, we start by noticing that if we take $\mathbf{x} = \Sigma^{-\frac{1}{2}} \mu$, $\mathbf{y} = \Sigma^{-\frac{1}{2}} \nu$ and $A = \Sigma^{-\frac{1}{2}} \Omega \Sigma^{-\frac{1}{2}}$, then

$$1 - \frac{\mu^\top \Omega^{-1} \nu}{\sqrt{\mu^\top \Sigma^{-1} \mu} \sqrt{\nu^\top \Omega^{-1} \Sigma \Omega^{-1} \nu}} = 1 - \frac{\mathbf{x}^\top A^{-1} \mathbf{y}}{\|\mathbf{x}\| \|A^{-1} \mathbf{y}\|} = \frac{1}{2} \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|} - \frac{A^{-1} \mathbf{y}}{\|A^{-1} \mathbf{y}\|} \right\|^2. \quad (6)$$

Now, for any two non-zero vectors $\mathbf{u}, \mathbf{v}$ we have the inequality

$$\left\| \frac{\mathbf{u}}{\|\mathbf{u}\|} - \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\|^2 \leq \frac{4\|\mathbf{u} - \mathbf{v}\|^2}{\|\mathbf{u}\|^2}.$$

Using this inequality with $\mathbf{u} = \mathbf{x}$, $\mathbf{v} = A^{-1} \mathbf{y}$ in (6) we have

$$1 - \frac{\mu^\top \Omega^{-1} \nu}{\sqrt{\mu^\top \Sigma^{-1} \mu} \sqrt{\nu^\top \Omega^{-1} \Sigma \Omega^{-1} \nu}} \leq 2 \frac{\|A^{-1} \mathbf{y} - \mathbf{x}\|^2}{\|\mathbf{x}\|^2} = \frac{\|\Sigma^{\frac{1}{2}} \Omega^{-1} \nu - \Sigma^{-\frac{1}{2}} \mu\|^2}{\mu^\top \Sigma^{-1} \mu}. \quad (7)$$

We shall first upper bound the norm in the numerator, by the triangle inequality and the definition of the operator norm $\|\cdot\|_{op}$, which we recall from above the statement of this theorem.

$$\begin{aligned} \|\Sigma^{\frac{1}{2}} \Omega^{-1} \nu - \Sigma^{-\frac{1}{2}} \mu\| &\leq \|\Sigma^{\frac{1}{2}} \Omega^{-1} \nu - \Sigma^{\frac{1}{2}} \Omega^{-1} \mu\| + \|\Sigma^{\frac{1}{2}} \Omega^{-1} \mu - \Sigma^{-\frac{1}{2}} \mu\| \\ &\leq \|\Sigma^{\frac{1}{2}}\|_{op} \|\Omega^{-1}\|_{op} \|\nu - \mu\| + \|\Sigma^{\frac{1}{2}}\|_{op} \|\Omega^{-1} - \Sigma^{-1}\|_{op} \|\mu\| \end{aligned} \quad (8)$$

Now, we apply the assumptions of the theorem to continue this bound. First, since $\Sigma - mI \succ 0$, we have $\|\Sigma^{-1}\|_{op} \leq \frac{1}{m}$. Then, since $\|\Sigma - \Omega\|_{op} < \frac{m}{2}$, we obtain $\frac{m}{2}I - (\Sigma - \Omega) \succ 0$. Adding the two matrix inequalities, we obtain that $\Omega - \frac{m}{2}I \succ 0$, which implies that $\|\Omega^{-1}\|_{op} \leq \frac{2}{m}$. Next, observe that

$$\|\Omega^{-1} - \Sigma^{-1}\|_{op} = \|\Sigma^{-1}(\Omega - \Sigma)\Omega^{-1}\|_{op} \leq \|\Sigma^{-1}\|_{op} \|\Omega - \Sigma\|_{op} \|\Omega^{-1}\|_{op} \leq \frac{2}{m^2} \|\Omega - \Sigma\|_{op}.$$

Putting these inequalities into (8) we have

$$\|\Sigma^{\frac{1}{2}} \Omega^{-1} \nu - \Sigma^{-\frac{1}{2}} \mu\| \leq \frac{2}{m} \|\Sigma^{\frac{1}{2}}\|_{op} \|\mu - \nu\| + \frac{2}{m^2} \|\Sigma^{\frac{1}{2}}\|_{op} \|\mu\| \|\Omega - \Sigma\|_{op}.$$

We plug the above into (7), and note that in (7), the denominator is lower bounded:

$$\mu^\top \Sigma^{-1} \mu = \|\Sigma^{-\frac{1}{2}} \mu\|^2 \geq \frac{\|\mu\|^2}{\|\Sigma^{\frac{1}{2}}\|_{op}^2}.$$

Thus, using the fact that $\|\mu\| \geq m$,

$$1 - \frac{\mu^\top \Omega^{-1} \nu}{\sqrt{\mu^\top \Sigma^{-1} \mu} \sqrt{\nu^\top \Omega^{-1} \Sigma \Omega^{-1} \nu}} \leq \frac{2\|\Sigma^{\frac{1}{2}}\|_{op}^4}{m^2} (\|\mu - \nu\| + \|\Omega - \Sigma\|_{op})^2.$$

This instantly concludes the proof in light of (4) and (5). $\square$

B Experiments

This appendix collects supporting material for the experiments in § 4. § B.1 details the construction of the historical measure used to form our initial view on returns, § B.2 describes the equity universe and data underlying both experiments, and § B.3 reports the full performance statistics and portfolio allocations summarized in Figure 2.

B.1 Historical Measure

Let $\mathbf{r}_i \in [-1, \infty)^n$, $i = 1, \dots, T$, denote the return vectors for n assets observed over a historical period of T days. For these observations, we take P to be the historical measure on $\mathbf{r}_1, \dots, \mathbf{r}_T$, defined for any subset $A \subset \mathbb{R}^n$ by

$$P(A) = \frac{1}{T} \sum_{i=1}^T \mathbf{1}_{\mathbf{r}_i \in A},$$

i.e., a uniform prior over the historical returns. If $\mathbf{r}$ denotes the random return vector under P , then $\mathbb{E}_P[\mathbf{r}] = \frac{1}{T} \sum_{i=1}^T \mathbf{r}_i$.

B.2 Equity Universe and Data

We restrict to firms with a closing price available on every trading date in the sample window from 1 January 2006 through 31 December 2025; since 1 January 2006 was a market holiday, the first observed trading date is 2 January 2006. This yields a common panel of 4,958 trading dates across the eligible firms. From this eligible pool, we rank firms by market capitalization as of 31 December 2023 and select the top 20.

The resulting universe consists of RELIANCE, TCS, HDFCBANK, AXISBANK, INFY, HINDUNILVR, BHARTIARTL, ITC, SBIN, LT, HCLTECH, KOTAKBANK, ICICIBANK, ASIANPAINT, NTPC, ADANIENT, SUNPHARMA, ULTRACEMCO, MARUTI, and TITAN.

This also serves as the Dataset 2 in Figure 1. However, Dataset 1 contains BHARTIARTL, KOTAKBANK, SBIN, MARUTI trained on data from 1 January 2020 to 31 December, 2023. Dataset 1 aims to serve as an illustration of the theoretical concavity result when the necessary conditions hold.

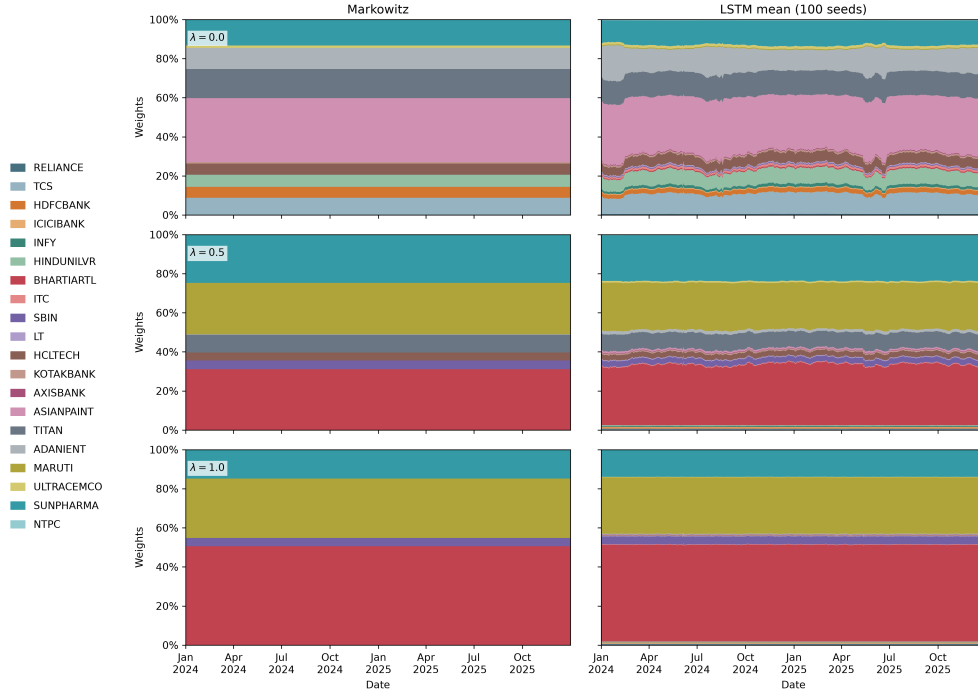
B.3 Portfolio Allocations and Performance Statistics

Figure 3 compares the Markowitz and mean LSTM allocations over 2024–2025 for three representative views, $\lambda = 0, 0.5$, and 1. The Markowitz allocation is constant by construction, while the LSTM allocation can vary from day to day. With the historical view ($\lambda = 0$), the LSTM makes substantial changes over time. As the view becomes more informative, these changes largely disappear and the LSTM converges toward an almost constant allocation that closely resembles the corresponding Markowitz portfolio. At $\lambda = 1$, both portfolios are concentrated primarily in BHARTIARTL, MARUTI, and SUNPHARMA.

Table 1 reports the full set of performance statistics underlying Figure 2 for the Indian equity universe described in the previous section, with LSTM entries averaged across 100 seeds and net metrics computed after 20 bps one-way transaction costs.

Table 1: Performance by λ .

λ	Model	Gross SR	Net SR	Sortino	Net Vol.	Net MDD	Final wealth	Daily turnover
0.0	Markowitz	0.050	0.019	0.029	13.3%	-25.0%	0.988	0.0081
h 0.0	LSTM	0.037	-0.021	-0.033	13.2%	-24.7%	0.978	0.0153
0.1	Markowitz	0.235	0.204	0.306	13.1%	-23.7%	1.037	0.0082
0.1	LSTM	0.220	0.150	0.229	13.0%	-23.1%	1.022	0.0180
0.2	Markowitz	0.670	0.638	0.949	13.0%	-21.4%	1.159	0.0082
0.2	LSTM	0.639	0.562	0.839	13.3%	-20.9%	1.139	0.0200
0.3	Markowitz	1.059	1.027	1.546	12.9%	-19.4%	1.279	0.0082
0.3	LSTM	0.959	0.885	1.304	13.5%	-19.1%	1.243	0.0197
0.4	Markowitz	1.456	1.426	2.172	13.3%	-16.9%	1.430	0.0079
0.4	LSTM	1.297	1.235	1.819	13.7%	-17.2%	1.371	0.0168
0.5	Markowitz	1.801	1.774	2.734	14.1%	-14.1%	1.609	0.0075
0.5	LSTM	1.733	1.691	2.580	14.0%	-14.4%	1.565	0.0118
0.6	Markowitz	1.902	1.877	2.905	14.7%	-13.4%	1.688	0.0073
0.6	LSTM	1.868	1.836	2.826	14.5%	-13.5%	1.659	0.0092
0.7	Markowitz	1.946	1.923	2.980	15.1%	-13.1%	1.738	0.0071
0.7	LSTM	1.908	1.881	2.904	14.9%	-13.2%	1.702	0.0082
0.8	Markowitz	1.953	1.931	2.992	15.4%	-13.4%	1.758	0.0069
0.8	LSTM	1.924	1.898	2.931	15.2%	-13.4%	1.728	0.0078
0.9	Markowitz	1.957	1.935	2.995	15.7%	-13.6%	1.777	0.0068
0.9	LSTM	1.932	1.907	2.940	15.5%	-13.6%	1.750	0.0077
1.0	Markowitz	1.957	1.936	2.987	16.0%	-13.8%	1.796	0.0066
1.0	LSTM	1.936	1.912	2.937	15.7%	-13.8%	1.769	0.0074

Figure 3: Portfolio allocations over 2024–2025 for $\lambda = 0, 0.5$, and 1 . The left column shows the Markowitz target allocations, while the right column shows the mean LSTM allocations across 100 random seeds.