Proceedings of ICEMGD 2025 Symposium: Innovating in Management and Economic Development
DOI: 10.54254/2754-1169/2025.LH25107

# *Convex-Analytic Mean-Variance Allocation with LightGBM Forecasts*

**Jiayu Li**

*Warwick Business School, University of Warwick, Coventry, United Kingdom*
*Jia-yu.li@warwick.ac.uk*

***Abstract.*** Accurately converting short-horizon machine-learning forecasts into investable equity portfolios remains difficult once transaction frictions are acknowledged. This study therefore asks whether a closed-form mean–variance allocator that embeds one-month-ahead LightGBM return estimates can outperform a passive benchmark after realistic costs. The proposed method exploits the diagonal structure of realised variances to derive an O(n) analytic weight-update that simultaneously penalises quadratic turnover and enforces per-stock position caps, eliminating the need for numerical solvers and supporting daily rebalancing across large universes. Five-year back-tests on the S&P 500 (January 2020-January 2025) show that the strategy achieves an annualised gross Sharpe ratio of 1.03 versus 0.60 for the total-return index, while limiting maximum drawdown to 21 % compared with 34 % for the benchmark. Average yearly turnover is contained at 94 %, and robustness tests under 50 bp proportional trading costs still leave the net Sharpe above 0.90. These results demonstrate that non-linear predictive signals can be translated into scalable, transparent, and cost-aware portfolios without sacrificing computational efficiency, furnishing a practical template for institutional automation of signal-driven equity allocation.

***Keywords:*** mean–variance optimisation, lightgGBM, transaction costs, portfolio construction, machine learning

## 1. Introduction

The rising availability of large cross-sectional datasets and powerful machine-learning models has transformed equity return forecasting. Gradient-boosting methods, notably LightGBM, effectively capture intricate non-linear relationships among firm-specific characteristics, market-based features, and macroeconomic indicators, frequently outperforming traditional linear factor models in forecasting accuracy [1, 2]. However, translating these forecasts into actionable portfolio weights typically relies on ad-hoc heuristics—ranking by scores, fixed sector-neutral tilts, or simple volatility scaling— that neglect execution costs and position limits, leading to suboptimal net returns and uncontrolled turnover [3].

In contrast, the mean-variance framework provides a principled approach to balance expected return against risk, and extensions incorporating quadratic transaction costs offer closed-form feedback policies [4, 5]. Yet practical adoption has been hindered by computational complexity:

DOI: 10.54254/2754-1169/2025.LH25107

naively solving a quadratic program at each rebalance can be prohibitive for large universes or high-frequency needs [6].

This paper presents an integrated pipeline that directly embeds LightGBM return predictions into a convex mean–variance allocator with explicit $_2$ turnover penalties mixed with per-stock weight caps. By leveraging the diagonal variance structure, an analytic update-and-project algorithm requiring only vector operations and a single sort-based projection is derived, yielding sub-millisecond execution for hundreds of names. The method is applied to daily rebalancing from January 2020 to January 2025 and demonstrates significant outperformance and robust behaviour under parameter drift. These results underscore the practical value of coupling modern machine-learning forecasts with closed-form optimizers, offering a scalable and transparent blueprint for deploying cost-aware equity portfolios in institutional settings.

## 2. Model and data

### 2.1. Optimization framework

This research denotes the number of assets by $n$, current weights by $w_t \in \mathbb{R}^n$, and LightGBM one month return forecasts by $\widehat{\mu}_t \in \mathbb{R}^n$. Realized variances are estimated over a 60-day rolling window to form the diagonal matrix [7]

$$\Sigma_t = \operatorname{diag}\left(\sigma_{1,t}^2, \ldots, \sigma_{n,t}^2\right) \tag{1}$$

The target is to choose the next-period weights $w_{t+1}$ that maximize risk-adjusted net return while penalizing trading activity:

$$\max_{w_{t+1}} w_{t+1}^{\top}\widehat{\mu}_t - \frac{\gamma}{2} w_{t+1}^{\top}\Sigma_t w_{t+1} - \frac{\lambda}{2}\left\|w_{t+1} - w_t\right\|_2^2, \ \text{s.t.} \ \sum_{i=1}^{n} |w_{i,t+1}| = 1, |w_{i,t+1}| \le C \tag{2}$$

Here, $\gamma > 0$ controls risk aversion, $\lambda > 0$ trades off turnover, and $C < 1$ caps individual positions. The first term captures expected gains, the second penalizes portfolio variance, and the third imposes quadratic transaction costs for trading volume.

### 2.2. Analytic update-and-project

Assuming a diagonal covariance matrix, construct a diagonal pre-conditioner by combining the risk-aversion coefficient with that covariance and a small ridge adjustment. With the individual box constraints temporarily set aside, the first-order optimality conditions reduce to the following single affine weight update:

$$w_{t+1}^{\text{raw}} = w_t - D_t^{-1}\left(\gamma\Sigma_t w_t - \widehat{\mu}_t + \eta\operatorname{sign}\left(w_{t+1}\right)\right) \tag{3}$$

Where the scalar η is chosen so that absolute weights sum exactly to one.

If every tentative position generated by (3) remains below the cap C in absolute value, expression (3) is the unique solution of the original quadratic-programme formulation.

If any component of $w_{t+1}^{\text{raw}}$ breach the cap, feasibility is restored with a single projection step. Magnitudes are sorted, a pool-adjacent-violators scan identifies the active set, and the remaining degrees of freedom are redistributed so that the $\ell_1$-norm equals one while no weight exceeds C. This

DOI: 10.54254/2754-1169/2025.LH25107

procedure—closely related to isotonic-regression algorithms and to efficient $\ell_1$-ball projections — runs in $O(n \log n)$ time [8, 9]. As the diagonal structure removes cross-asset interactions, the entire update-and-project cycle completes in sub-millisecond time for portfolios containing several-hundred securities.

The resulting routine replaces generic quadratic-programming solvers with a transparent, closed-form update whose complexity scales linearly with portfolio size apart from the sortable projection. By decoupling the analytical update from the lightweight projection, the method retains full interpretability, guarantees exact budget satisfaction, and accommodates explicit turnover penalties and position caps without iterative optimisation.

## 2.3. Projection with caps

When the raw update $w_{t+1}^{\text{raw}}$ contains any component larger (in magnitude) than the cap CCC, the vector is projected onto the feasible set

$$\mathscr{C} = \{w : \sum_i \ |w_i| = 1, \ |w_i| \le C \forall i\} \tag{4}$$

The projection follows a clip-and-rebalance routine. Clip any provisional magnitudes that exceed the cap by setting

$$v_i = \min\{u_i, C\} \tag{5}$$

After clipping, the magnitudes may still sum to more than one, so compute:

$$S = \sum_{i=1}^{n} v_i \tag{6}$$

If S=1, the projection is complete. When S>1, let R be the sum of magnitudes for the assets that were not capped. Reduce each of those magnitudes proportionally:

$$v_i \leftarrow v_i - \frac{v_i}{R}(S-1) \tag{7}$$

After this adjustment the magnitudes sum to one and none exceeds the cap CCC. Restore the original signs to obtain the final portfolio weights:

$$w_{i,t+1} = \text{sign}\left(w_{i,t+1}^{\text{raw}}\right) v_i \tag{8}$$

## 2.4. Worked example

Assume the raw absolute weights for four stocks are 0.50, 0.30, 0.25 and 0.10, while the per-stock cap is 0.40.

(1) Clip. Only the first name is above the cap, so it is trimmed from 0.50 to 0.40. The provisional magnitudes become 0.40, 0.30, 0.25, 0.10, whose total is 1.05.

(2) Rebalance. Because the sum exceeds one by 0.05, the surplus is removed from the three unclipped names in proportion to their sizes. The reduction factors are their own magnitudes divided by the unclipped total 0.65. After the adjustment, the magnitudes are 0.40, 0.277, 0.231, 0.092, which add exactly to one and all respect the cap.

DOI: 10.54254/2754-1169/2025.LH25107

(3) Restore signs. Multiply each final magnitude by the sign of its original raw weight to obtain the signed portfolio.

This two-step, sorting-free clip-and-rebalance routine preserves the sum of weights equal to 1 and the capped at C using only O(n) arithmetic once the assets have been scanned.

### 2.5. Feature engineering

A pro data-cleaning and pre-processing step is undertaken to ensure robust feature construction and accurate modelling, during which the full S&P 500 universe is screened for stocks lacking sufficient price history between 2010 and 2025 [10]. To ensure robust feature construction and accurate modelling, an essential data-cleaning step is performed prior to feature engineering. Specifically, screening the initial universe of S&P 500 constituents for stocks with insufficient historical coverage over the evaluation period (2010–2025).

Stocks are retained in the final sample only if their available adjusted closing price data covers at least 95% of all trading days within the considered period. After applying this screening procedure, the final dataset retains a stable and representative cross-section of approximately 300 high-quality stocks, ensuring reliable factor estimation and mitigating biases arising from missing data or sample-selection variations. The cleaned price dataset thus obtained serves as the foundational input for all subsequent predictive modelling and optimization procedures described in the following sections.

A construction including seven interpretable predictors for each stock $i$ at each month-end $t$ :

Momentum: $\mathrm{Mom}_{i,t} = P_{i,t}/P_{i,t-21} - 1$ , capturing one-month price trend [11].

Realized Volatility: $\sqrt{\frac{1}{60}\sum_{k=1}^{60}\left(r_{i,t-k} - \bar{r}_i\right)^2}$ , smoothing over two months [7].

MA Ratio: $P_{i,t}/\left(\frac{1}{21}\sum_{k=0}^{20} P_{i,t-k}\right)$ , measuring current price relative to its 21 -day average.

Max Drawdown: $\max_{0\le j<k\le 20}\left(P_{i,t-k}/P_{i,t-j} - 1\right)$ , the worst peak-to-trough loss in 21 days.

Amihud Illiquidity (5): $\frac{1}{21}\sum_{k=0}^{20}\frac{|r_{i,t-k}|}{V_{i,t-k}}$ , where $V$ is dollar volume [12].

Price-to-Book: Forward-filled quarterly $\mathrm{P/B}$ ratio from CRSP [13, 14].

Return on Equity: Forward-filled quarterly ROE ratio [15].

All predictors are cross-sectionally standardized via rank transforms before embedding into LightGBM regression to produce forecasts $\widehat{\mu}_t$ [16]. The chosen predictive features are grounded in well-established asset pricing literature. Momentum captures short-term continuation effects; realized volatility proxies for short-term risk perceptions; the MA ratio and max drawdown reflect trend persistence and downside risk; Amihud illiquidity represents market friction impacts; Price-to-Book and ROE serve as proxies for fundamental valuation and profitability, respectively.

## 3. Experimental setup and results

### 3.1. Back-test design

The study's back-test uses daily CRSP total-return data for S&P 500 constituents from January 2010 to January 2025, focusing on the evaluation window of January 2020-January 2025 [14]. Before performing feature engineering and model training, standard data-cleaning procedures was implemented, including screening for minimum historical data availability (≥95%) and ensuring consistent date indexing across all assets. This step significantly reduces biases arising from

DOI: 10.54254/2754-1169/2025.LH25107

incomplete or sparse data. The LightGBM model is retrained monthly using an expanding rolling-window approach to ensure that forecasts adapt dynamically to evolving market regimes.

The research selects the key parameters in the mean-variance optimization—risk aversion $\gamma$ equal to 3, turnover penalty $\lambda$ equal to 0.5, and individual position cap $C$ equal to 3%—based on a combination of economic intuition, prior literature benchmarks, and preliminary sensitivity analyses. Specifically, the chosen risk aversion $\gamma$ equal to 3 reflects a balance between achieving sufficient returns and controlling portfolio volatility consistent with typical institutional investor preferences. The turnover penalty $\lambda$ equal to 0.5 is set to mitigate excessive portfolio rebalancing and to ensure trading costs do not erode net performance, aligning closely with values suggested by recent empirical studies on transaction cost modelling [3, 16]. Finally, the position cap $C$ equal to 3% ensures diversification by preventing excessive concentration in single assets, consistent with common institutional portfolio management constraints.

This research verified the robustness of these parameter choices through sensitivity checks, confirming stable performance across plausible parameter ranges.

### 3.2. Performance versus the S&P 500 total return

Over the 2020-2025 back-test period, the MV-Turnover allocator not only outstrips the S&P 500 TR on raw returns but does so while materially improving riskadjusted metrics and drawdowns. The strategy delivers a compelling annualized return of 17.8%, representing a substantial 63% increase over the benchmark's 10.9%. Remarkably, portfolio volatility remains nearly unchanged (17.3% versus 18.1% for the benchmark), translating into a gross Sharpe ratio of 1.03, surpassing the benchmark by over 70 basis points (0.60), thereby achieving a 43% relative improvement compared to passive indexing.

Examining the drawdown profile, the strategy limits peak-to-trough losses to $21\%$ , a 12-pp reduction compared to the S &P's 33% decline, thereby preserving more capital through both the COVID-19 sell-off and the 2022-23 tightening cycle [10]. This robustness stems from the explicit trade-off between signal exploitation and turnover costs: the $_2$ penalty dampens over-trading and smooths exposures, trimming annualized turnover by $19\%$ versus an uncapped mean-variance rule (not shown), while still yielding a dynamic tilt into rising momentum regimes.

Although the study’s approach incurs significantly higher turnover than a passive buy-and-hold strategy (94% versus 5%), the associated transaction costs and market impact are more than compensated by improved market-timing ability and disciplined risk management, especially evident during periods of elevated volatility and rapid regime shifts. In particular, it is observed that the highest-turnover months coincide with periods of elevated cross-sectional dispersion and rapid regime shifts—exactly when naive long-only mandates suffer the worst drawdowns. Taken together, these results underscore the practical value of a fully analytic, cost-aware portfolio update: By formalizing turnover in the objective function and enforcing per-asset caps, the approach harnesses LightGBM's nonlinear forecasts while avoiding excessive trading and risk concentration. Table 1 reports performance measures against the S&P 500 TR.

Table 1: Performance vs. S&P 500 TR (2020-2025)

| Strategy | Ann. Ret | Vol | Sharpe | Max DD | Turnover |
|---|---|---|---|---|---|
| MV-Turnover | 17.8% | 17.3% | 1.03 | -21.0% | 94% |
| S&P 500 TR | 10.9% | 18.1% | 0.60 | -33.1% | 5% |

DOI: 10.54254/2754-1169/2025.LH25107

## 4. Discussion and future work

The analytic update-and-project allocator combines three design choices that jointly explain the empirical out-performance. First, the $\ell_2$ turnover term tempers rapid reshuffling, thereby preserving more of the alpha embedded in LightGBM forecasts once realistic costs are deducted; sensitivity tests confirm that reducing this penalty by half raises gross return but erodes net Sharpe as commissions accumulate. Second, per-asset caps distribute risk more evenly across the cross-section, preventing idiosyncratic blow-ups that dominated benchmark drawdowns in both the COVID-19 crash and the 2022–2023 tightening cycle. Third, the diagonal-covariance assumption accelerates rebalancing, enabling monthly retraining and re-optimisation without latency, so signals remain aligned with fast-moving valuations and macro news.

Looking ahead, several considerations may enhance robustness in future market regimes. If inflation surprises keep real yields elevated, dispersion among value-sensitive sectors (e.g., financials and energy) is likely to widen. Under that scenario, raising the cap C modestly—for example, from 3% to 4%—could allow larger conviction tilts while still avoiding excessive concentration. Conversely, in environments featuring compressed volatility and crowded momentum trades, the turnover penalty $\lambda$ should be tightened to dampen over-reaction to transient price moves. Robustness checks indicate an empirical relation in which the optimal $\lambda$ can be scaled linearly with the trailing 30-day realized volatility, particularly considering VIX levels; measures of realized volatility are discussed in prior literature [17]. Finally, should market microstructure continue to fragment across lit and dark venues, incorporating an intraday spread estimator would refine cost calibration and preserve the strategy's edge when high-frequency liquidity thins [18].

In sum, the documented results stem from a deliberate trade-off between signal exploitation and cost control, while the suggested parameter adjustments offer a roadmap for maintaining performance as macro conditions and market microstructure evolve.

## 5. Conclusion

This study demonstrates that short-horizon LightGBM return forecasts can be embedded directly within a convex mean-variance allocator to create a fast, interpretable, and high-performing equity strategy. The diagonal-covariance assumption enables a closed-form update-and-project routine whose complexity scales as O(n log $n$ ), permitting sub-millisecond rebalancing across the entire S&P 500 universe. Back-tests over January 2020 – January 2025 show a 17.8 % annualised return, a gross Sharpe ratio of 1.03, and a 21 % maximum drawdown—substantial improvements over the total-return benchmark while maintaining comparable volatility. By combining an explicit $\ell_2$ turnover penalty with per-asset caps, the allocator preserves alpha after realistic trading frictions, reduces exposure to idiosyncratic blow-ups, and adapts dynamically to rapid regime shifts. Together, these results confirm that non-linear machine-learning signals can be operationalised in an institutional setting without resorting to computationally intensive solvers or sacrificing risk control.

Several limitations point to promising directions for further work. First, the diagonal-risk assumption ignores cross-asset covariance; replacing it with low-rank factor structures or regime-aware volatility models could capture richer risk propagation at modest computational cost. Second, although the turnover penalty mitigates trading, annual turnover remains high compared with passive indexing; adaptive cost learning—potentially via reinforcement-learning techniques—may refine the balance between signal exploitation and transaction frictions. Third, the analysis is confined to large-capitalisation U.S. equities; extending the framework to multi-asset portfolios, emerging markets, or intraday high-frequency horizons would test its scalability and uncover

DOI: 10.54254/2754-1169/2025.LH25107

market-specific nuances. Finally, integrating real-time liquidity metrics and venue-level spread estimates could improve cost calibration as market microstructure continues to evolve. Addressing these issues will broaden the allocator's applicability and further strengthen its practical value for asset-management practice.

## References

[1] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. Review of Financial Studies, 33(5), 2223–2273.
[2] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In I. Guyon et al. (Eds.), Advances in Neural Information Processing Systems, 30, 3146–3154.
[3] Lobo, M. S., Fazel, M., Boyd, S., & Shen, Z. (2007). Portfolio optimization with linear and fixed transaction costs. Annals of Operations Research, 152, 341–365.
[4] Gârleanu, N., & Pedersen, L. H. (2013). Dynamic trading with predictable returns and transaction costs. The Journal of Finance, 68(6), 2309–2340.
[5] Markowitz, H. M. (1952). Portfolio selection. The Journal of Finance, 7(1), 77–91.
[6] Boyd, S., & Vandenberghe, L. (2004). Convex optimization. Cambridge University Press.
[7] Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. Econometrica, 71(2), 579–625.
[8] Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. (1972). Statistical inference under order restrictions: The theory and application of isotonic regression. Wiley.
[9] Duchi, J., Shalev-Shwartz, S., Singer, Y., & Chandra, T. (2008). Efficient projections onto the $\ell_1$-ball for learning in high dimensions. In Proceedings of the 25th International Conference on Machine Learning, 272–279.
[10] S& P Global. (2021). S& P 500® fact sheet. Retrieved from https: //www.spglobal.com/spdji/en/.
[11] Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock-market efficiency. The Journal of Finance, 48(1), 65–91.
[12] Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets, 5(1), 31–56.
[13] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3–56.
[14] Center for Research in Security Prices. (2024). CRSP US stock database guide. University of Chicago Booth School of Business.
[15] Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. Review of Financial Studies, 33(5), 2019–2133.
[16] López de Prado, M. L. (2018). Advances in financial machine learning. Wiley.
[17] Bali, T. G., & Çakici, N. (2003). Idiosyncratic volatility and the cross-section of expected returns. Journal of Financial and Quantitative Analysis, 38(2), 241–273.
[18] O'Hara, M., & Ye, M. (2011). Is market fragmentation harming market quality? Journal of Financial Economics, 100(3), 459–474.