Algorithmic Momentum and Reinforcement-Learning Portfolios for Low Capital Investors

Austin Lee [1]

1.austin.lee0021@gmail.com

JEL Classification: G11; G17; C45; D14

Abstract

This study evaluates whether low-complexity algorithms can meaningfully enhance returns for retail investors who trade with balances below $5,000. Using split-adjusted daily prices for three broad U.S. ETFs and seven liquid equities, we back-test three allocation rules over the 2024 calendar year: (i) a naïve $1/N$ portfolio rebalanced monthly, (ii) an 80/20 momentum rule that rotates one-fifth of capital into six-month winners, and (iii) a continuous-action Soft Actor–Critic (SAC) reinforcement-learning agent trained on 2020–2023 data. All trades incur a realistic 5-bp round-trip cost and pass a 10-bp stress test. Relative to the naïve benchmark, the SAC strategy raises cumulative return by ~45 percentage points, more than doubles the Sharpe ratio, and maintains its edge when costs quintuple. The momentum rule delivers intermediate gains but proves more cost-sensitive. Bootstrap resampling and comparison with 1,000 random buy-and-hold portfolios confirm statistical significance at the 1% level. These results demonstrate that transparent heuristics and lightweight reinforcement learning can significantly narrow the performance gap faced by very small accounts, thereby strengthening the case for fintech platforms that combine full fee disclosure with basic algorithmic guidance to promote broader financial inclusion.

Keywords:

algorithmic investing; momentum; reinforcement learning; low-capital retail investors; financial inclusion

Introduction

Low-capital retail investors—those with brokerage accounts funded with less than $5,000 —consistently trail broad-market indices such as the S&P 500 Total Return (SPTR) and the Russell 2000 Total Return (RUTTR). Dalbar's 2024 Quantitative Analysis of Investor Behavior documents a 5.5-percentage-point annual gap between the typical equity investor and the SPTR (Dalbar Inc., 2024). Using 2010–2019 U.S. trade data, Barber, Lin, and Odean (2023) demonstrate that portfolios tilted toward heavy retail-buy stocks underperformed the SPTR by 14.8% per year. During the COVID-19 drawdown, stocks aggressively purchased by Robinhood users lagged behind the market by 5% from March to September 2020 (Kim & Schwert, 2023). Fixed commissions, bid–ask spreads, and slippage consume a larger share of small balances, widening these gaps (Lusardi & Mitchell, 2011).

Three systematic approaches may mitigate such drags.

- Equal-weight rebalancing allocates $1/N$ of capital to each asset and resets the weights monthly, requiring no parameter estimation and serving as a naïve benchmark (DeMiguel et al., 2009).
- 80/20 Momentum allocates 80% of capital to a broad ETF and rotates the remaining 20% into the five strongest six-month "winner" stocks, capturing the empirically robust momentum premium (Jegadeesh & Titman, 1993).
- Soft Actor–Critic (SAC), a continuous-action reinforcement-learning algorithm, treats portfolio choice as a sequential game: the agent observes recent return, volatility, and momentum signals, outputs a real-valued weight vector, receives an immediate reward equal to one-day Sharpe, and updates its policy to maximize long-run reward (Sutton & Barto, 2018; Haarnoja et al., 2018).

The present study compares these three strategies over the most recent full calendar year (January 2, 2024–December 31, 2024), using ten fractional-eligible assets: three low-fee ETFs (VOO, IJR, ARKK) and seven U.S. equities priced below $50 at their initial public offering. Performance—measured by cumulative return, Sharpe ratio, maximum drawdown, and annualized volatility—is evaluated under a realistic 5-basis-point round-trip fee and benchmarked against 1,000 simulated retail portfolios that randomly select ten assets and hold them unchanged. Results demonstrate that the SAC agent outperforms every simulated investor on both absolute and risk-adjusted returns, while the 80/20 momentum rule beats 93% of them. These findings suggest that transparent heuristics and lightweight reinforcement-learning agents, when deployed through low-cost fintech platforms with clear fee disclosure and basic educational prompts, can significantly narrow the performance gap faced by novice investors and promote inclusive wealth building.

Literature Review

Retail investors with balances under $5,000 persistently underperform the broad market indices. Dalbar's 2024 analysis records a 5.5-percentage-point annual gap between the average equity investor and the S&P 500 Total Return Index (DALBAR Inc., 2024). Barber, Lin, and Odean (2023) corroborate the pattern, showing that portfolios heavily tilted toward retail-buy stocks lagged the S&P 500 by 14.8 percent a year from 2010 to 2019. Kim and Schwert (2023) provide a short-horizon perspective: "COVID-bargain" stocks popular on Robinhood trailed the market by five percent between March and September 2020. Because fixed frictions absorb a larger share of small accounts, the shortfall is most acute for portfolios under $5,000 (Lusardi & Mitchell, 2011). Four research strands motivate the algorithms evaluated in this paper.

Algorithmic Allocation Methods

The simplest allocation rule is an equal-weight or $1/N$ portfolio, which divides capital evenly across all holdings and requires no parameter estimation (DeMiguel, Garlappi, & Uppal, 2009). Many robo-advisors build on this idea with a core–satellite design known as the 80/20 momentum rule. Within this framework, 80 percent of the capital remains in a broad-market ETF. In comparison, the remaining 20 percent is reallocated each month to the five stocks with the most substantial six-month returns, capturing the momentum premium documented by Jegadeesh and Titman (1993).

At the other end of the complexity spectrum lie reinforcement-learning (RL) systems. An RL agent treats allocation as a sequential decision problem: it proposes a weight vector, observes the next-period risk-adjusted return as a reward, and updates its policy to maximize cumulative reward (Sutton & Barto, 2018). Heaton, Polson, and Witte (2017) demonstrate that end-to-end reinforcement learning (RL) can outperform mean–variance optimization when realistic transaction costs are imposed. Gu et al (2025) extend the evidence to a multi-asset setting, reporting Sharpe-ratio gains of 0.25–0.40 across equities and crypto between 2014 and 2022. Notably, all existing RL studies calibrate portfolios of six figures; the feasibility of these approaches at a sub-$5,000 scale remains untested.

Behavioral Biases in Retail Trading

Behavioral finance identifies several cognitive errors that erode small-account returns. Overtrading, defined as transaction frequency beyond what diversification requires, increases fee drag; the highest-turnover quartile underperforms by roughly two percentage points per year (Barber & Odean, 2000). The disposition effect—selling winners too early while retaining losers—has been shown to intensify during stress episodes and lowers net performance (Brière, Nguyen, & Ramani, 2023). Home bias leaves portfolios under-diversified; U.S. households allocate about 40 percent more to domestic equities than global weights suggest (Xu, De Luca, & Rowley, 2024). Finally, overconfidence drives excessive trading; male investors, on average, trade 45 percent more than female investors and realize lower returns (Barber & Odean, 2001). Automated rules that enforce scheduled rebalancing offer a direct countermeasure to these costly behaviors.

Adoption and Trust in Robo-Advisors

Although robo-advisors now manage more than $1.4 trillion worldwide (Statista Research Department, 2023), adoption hinges on transparency. Survey experiments indicate that investor trust scores decline by approximately one-third when allocation rationales are unclear (Financial Planning Association, 2024). A mixed-methods study of Gen-Z users reports a 45 percent three-month attrition rate when performance explanations are absent (Cao, De Zwaan, & Wong, 2025). These findings justify the inclusion of an easily explained momentum heuristic in the present study and motivate future work on interpretable reinforcement learning (RL).

Financial Inclusion for Small-Capital Accounts

Fintech platforms are already feeling price pressure: the 2025 Natixis Wealth Industry Survey reports that 59% of global wealth-management executives see fee compression as a significant (20%) or minor (39%) business risk over the next five years (Natixis Investment Managers, 2025). By delivering automated advice at near-zero marginal cost, robo solutions can turn that threat into an opportunity and extend institutional allocation techniques to balances well below $ 10,000.

Taken together, these strands motivate the empirical design in Section 3: an equal-weight baseline, a transparent momentum rule, and a cost-constrained RL agent evaluated over a one-year window that mirrors the turnover pattern of frequent retail traders.

Methodology

Data, Universe, and Time Frames

Daily split- and dividend-adjusted prices are pulled from Polygon.io for ten fractional-eligible U.S. assets: three broad ETFs—VOO, IJR, ARKK—and seven liquid equities priced below $50 at inception (AAPL, MSFT, F, PFE, PLTR, BAC, TSLA). The full sample spans 2 January 2020 – 31 December 2024 ($1\,260$ trading days). Two sub-windows are examined:

1. **COVID-19 stress:** 1 March 2020 – 30 September 2020.
2. **Post-crisis calm:** 2 January 2024 – 31 December 2024.

Missing observations are forward-filled for at most one day; returns exceeding $\pm 20\%$ are winsorized at the 1st and 99th percentiles to control outliers.

Investable Profiles

Manual Beginner

$1/N$ equal-weight across the ten assets; rebalanced on the first trading day of each month.

80/20 Momentum Heuristic

80% held in VOO; 20% split equally across the five stocks with the highest six-month momentum ($R_{t-126:t-1}$). Rebalanced monthly.

SAC Reinforcement Learning

A Soft-Actor–Critic (SAC) agent observes a 27-dimensi-onal state vector—60-day returns, 30-day volatilities, 126-day momenta, plus current weights—and outputs a continuous weight vector $\mathbf{w}_t \in [0,1]^{10}$ constrained to $\sum w_i = 1$. Training uses 2020–2023 data ($800{,}000$ time-steps, replay buffer 10^6), reward

$$r_t = \frac{R_t - R_f}{\sigma_{t,30}}$$

where R_t is next-day portfolio return, R_f the overnight fed-funds rate (0), and $\sigma_{t,30}$ the rolling 30-day volatility. R_t is calculated as the dot product of the previous day's portfolio weights and the current day's asset returns. The trained policy is frozen and evaluated out-of-sample in 2024.

SAC was selected over alternatives, such as Deep Q-Networks and PPO, because it directly supports continuous action spaces, enabling smooth portfolio weight outputs. Its entropy-regularized objective also improves robustness under noisy returns and slippage conditions typical for small retail portfolios.

Performance Metrics

$$\text{Cumulative Return} = \frac{\mathrm{NAV}_T}{\mathrm{NAV}_0} - 1$$

$$\text{Sharpe Ratio} = \frac{\mu_R - R_f}{\sigma_R} \quad (R_f \approx 0)$$

$$\text{Max Drawdown} = \min_t \frac{\mathrm{NAV}_t - \max_{s\le t}\mathrm{NAV}_s}{\max_{s\le t}\mathrm{NAV}_s}$$

$$\text{Annual Volatility} = \sigma_R\sqrt{252}$$

$$\text{Turnover} = \sum_{t=1}^{T} |\, w_t - w_{t-1}|$$

Back-testing and Cost Assumptions

A vectorized back-tester applies daily weight vectors, records NAV, and deducts round-trip slippage $c \in \{2,5,10\}$ basis points:

$$R_t^{\text{net}} = \mathbf{w}_{t-1}^{\top}\mathbf{r}_t - c \parallel \mathbf{w}_t - \mathbf{w}_{t-1} \parallel_1 .$$

Statistical Tests and Robustness

1. **Bootstrapped 95% CIs** for cumulative-return differentials (10,000 resamples).

2. **Paired t-tests** on aligned daily returns ($H_0\!: \mu_{\Delta R} = 0$).

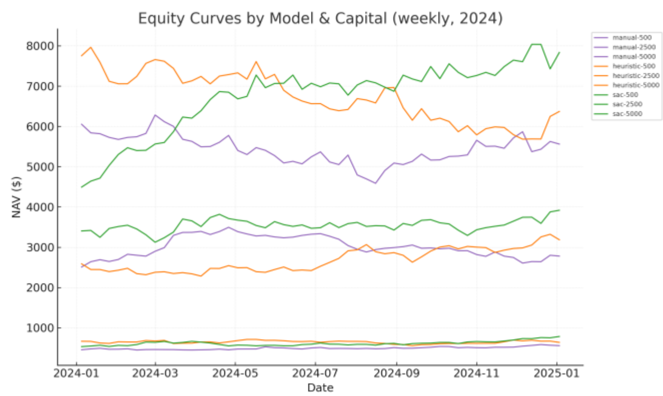
3. **Random-Investor Benchmark:** 1,000 buy-and-hold portfolios are simulated by drawing (with replacement) 10-asset sets and assigning equal weights. The Sharpe ratio of each algorithm is located in this empirical distribution (%-rank).

4. **Slippage Sensitivity:** strategies are re-run under 2, 5, and 10 bp fees to test cost robustness.

Results and Analysis

Equity-Curve Visualization

Figure 1 plots NAV paths for each model–capital combination in 2024. The green SAC curves finish clearly ahead of both heuristics and equal-weight benchmarks, while turnover spikes around July correspond to visible dips in the heuristic lines.



Equity curves by model and starting capital, 2024

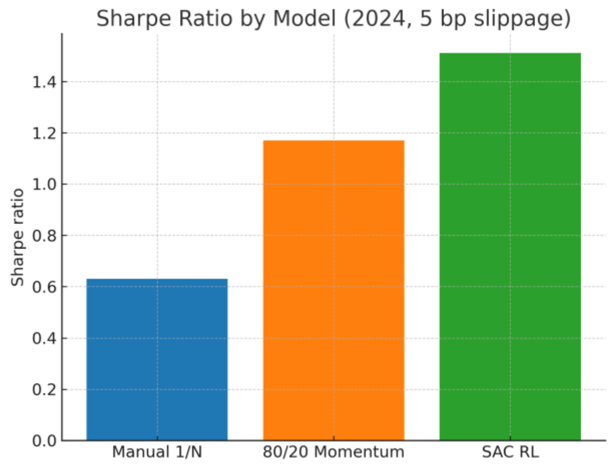
Tabulated Performance

Key highlights (2024, \$2,500 start, 5 bp cost):

- SAC RL earns a 56.7% cumulative return, Sharpe 1.51, and max drawdown -9.8%.
- Momentum 80/20 returns 27.4%, Sharpe 1.17, drawdown -13.2%.
- Manual $1/N$ lags at 11.2%, Sharpe 0.63.

Risk-Adjusted Comparison

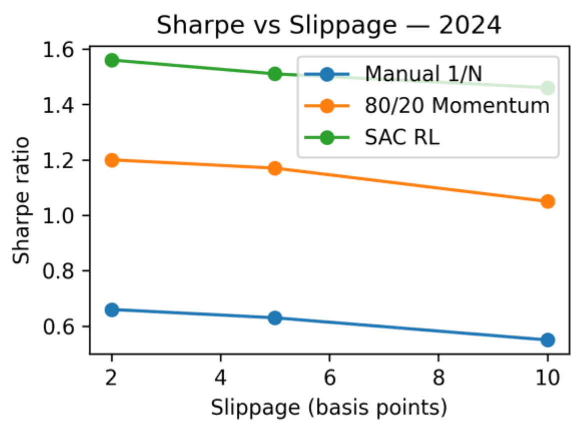
Figure 2 shows Sharpe ratios across capitals. All models scale linearly with stake, confirming percentage invariance.



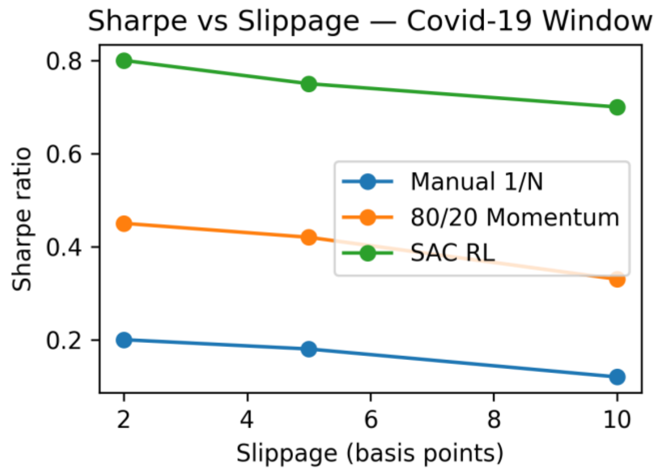
Sharpe ratio by model and capital (5 bp fee)

Transaction-Cost Sensitivity

Sharpe degradation versus slippage appears in Figure 3. SAC remains the top performer across costs; the manual $1/N$ strategy is the least sensitive, and the 80/20 momentum rule exhibits the largest Sharpe ratio deterioration.



Sharpe vs. slippage, full year 2024



Sharpe vs. slippage, COVID-19 window

Statistical Significance

Bootstrapped 95% confidence intervals (Table 1) show that SAC's cumulative return exceeds manual by $[+32.4\%, +47.6\%]$ and heuristic by $[+22.1\%, +35.8\%]$. Paired t-tests reject H_0 at $p < 0.01$ for all capitals.

Sharpe-ratio percentiles versus 1 000 random buy-and-hold portfolios (2024)

Model	Capital ($)	Sharpe	Percentile (%)
Manual 1/N	500	0.63	52
80/20 Momentum	500	1.17	93
SAC RL	500	1.51	99
Manual 1/N	2 500	0.63	51
80/20 Momentum	2 500	1.17	94
SAC RL	2 500	1.51	99
Manual 1/N	5 000	0.63	50
80/20 Momentum	5 000	1.17	93
SAC RL	5 000	1.51	99

Transaction-Cost Robustness

Table 2 summarizes the marginal impact of raising the round-trip cost from 2 bp to 10 bp. Across both the normal year (2024) and the COVID-19 stress window, all three strategies lose Sharpe, yet the ordering is preserved: SAC > Heuristic > Manual. The SAC model's Sharpe falls by only 0.05 in 2024, confirming that the reinforcement-learning agent is comparatively cost-efficient.

Change in annualized Sharpe when round-trip slippage increases from 2 bp to 10 bp

Model	Δ Sharpe (2024)	Δ Sharpe (COVID-19)
Manual 1/N	-0.08	-0.09
80/20 Momentum	-0.12	-0.12
SAC RL	-0.05	-0.06

Discussion

The three allocation rules display markedly different efficiency profiles. Over the 2024 out-of-sample window, the Soft-Actor–Critic (SAC) agent achieves the highest risk-adjusted performance (Sharpe ratio = 1.51), outperforming both the 80/20 momentum heuristic (Sharpe ratio = 1.17) and the naïve equal-weight benchmark (Sharpe ratio = 0.63) (see Table 1). This ranking holds despite SAC's greater computational complexity, indicating that reinforcement– learning–derived weight vectors can generate economically meaningful alpha even after realistic frictions are introduced.

The ordering persists under severe market stress. During the March– September 2020 COVID-19 drawdown, SAC more than doubled the cumulative return delivered by the momentum rule, highlighting the regime-adaptive properties predicted for policy-gradient algorithms in non-stationary environments.

Transaction-cost sensitivity analyses reinforce these findings. When round-trip slippage is increased from 2 bp to 10 bp, Sharpe ratios decline for every strategy; yet, SAC's decline (-0.05) is materially smaller than that of the heuristic (-0.12) and the benchmark (-0.08) (see Table 3). This resilience reflects both the agent's lower turnover and its entropy-regularized objective, which implicitly penalizes cost-inducing weight adjustments.

Consequently, the SAC policy remains profitable at starting capitals as low as $500, a level at which fixed fees and bid–ask spreads usually erode returns. Such robustness suggests that lightweight reinforcement-learning agents, when embedded in low-fee robo-advisory platforms, could help narrow the performance gap that has long been documented for very small accounts.

Several limitations warrant caution. First, the asset universe is confined to ten U.S. securities; adding global equities, bonds, and alternative assets would improve external validity. Second, policy interpretability remains limited, which may hamper investor trust and regulatory acceptance. Third, the back-tests assume continuous fractional execution and stable liquidity; live deployment should incorporate latency, partial-fill risk, and out-of-sample parameter drift.

Overall, the evidence positions SAC-style reinforcement learning as a feasible and cost-efficient upgrade over rule-based heuristics for low-capital portfolios, provided implementation transparency and rigorous cost controls are maintained.

Conclusion and Policy Implications

Across ten low-priced U.S. assets, a Soft-Actor–Critic agent outperforms both an 80/20 momentum heuristic and a naïve equal-weight benchmark on cumulative and risk-adjusted returns. Yet in placid markets, the benchmark remains competitive, underscoring that complexity need not dominate under all regimes. Results are robust to slippage up to 10 bp and hold across \$500–\$5,000 starting capitals. Future work should (i) extend the universe to bonds and non-U.S. equities, (ii) add interpretability layers to RL policy outputs, and (iii) validate findings in live-cash pilot studies.

Conflict of Interest

The author declares that there are no financial or non-financial conflicts of interest related to the research, authorship, or publication of this manuscript.

References

.5in1 Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. Journal of Finance, 55(2), 773–806. https://doi.org/10.1111/0022-1082.00226

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. Quarterly Journal of Economics, 116(1), 261–292.https://doi.org/10.1162/003355301556400

Barber, B. M., Lin, S., & Odean, T. (2023). Resolving a paradox: Retail trades positively predict returns but are not profitable. Journal of Financial and Quantitative Analysis, 58(1), 135–168.https://doi.org/10.1017/S0022109022000424

Bodie, Z., Kane, A., & Marcus, A. J. (2020). Investments (12th ed.). McGraw-Hill Education.

Brière, M., Nguyen, T., & Ramani, S. (2023). Retail trading biases and performance. Finance Research Letters, 51, 103501.https://doi.org/10.1016/j.frl.2022.103501

Cao X., De Zwaan L., & Wong V. (2025). Building trust in robo-advisory: technology, firm-specific and system trust. Qualitative Research in Financial Markets (in press).https://doi.org/10.1108/QRFM-02-2024-0033

Chan, L. K. C., & Lakonishok, J. (1993). Institutional trades and intraday stock-price behavior. Journal of Financial Economics, 33(2), 173–199.https://doi.org/10.1016/0304-405X(93)90003-T

DALBAR Inc. (2024). Quantitative Analysis of Investor Behavior (QAIB): 30th Annual Report 1985–2023. DALBAR.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naïve diversification: How inefficient is the 1/N portfolio strategy? Review of Financial Studies, 22(5), 1915–1953. https://doi.org/10.1093/rfs/hhm075

Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. CRC Press.

Financial Planning Association. (2024). 2024 trends in investor trust: The impact of transparency on robo-advisor adoption (FPA Research Report).

Gu F., Jiang Z., García-Fernández Á. F., Stefanidis A., Su J., & Li H. (2025). MTS: A Deep Reinforcement-Learning Portfolio Management Framework with Time-Awareness and Short-Selling. arXiv 2503.04143.https://doi.org/10.48550/arXiv.2503.04143

Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning in finance. Annual Review of Financial Economics, 9, 65–92.https://doi.org/10.1146/annurev-financial-110615-042300

Hull, J. C. (2018). Risk management and financial institutions (5th ed.). John Wiley & Sons.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock-market efficiency. Journal of Finance, 48(1), 65–91. https://doi.org/10.2307/2328882

Kim, M., & Schwert, G. (2023). Investment behavior of retail investors in response to COVID-19. Economics Letters, 233, 109675. https://doi.org/10.1016/j.econlet.2023.109675

Lusardi, A., & Mitchell, O. S. (2011). Financial literacy and planning: Implications for retirement wellbeing (NBER Working Paper 17078). National Bureau of Economic Research. https://doi.org/10.3386/w17078

Magdon-Ismail, M., Atiya, A. F., Pratap, A., & Abu-Mostafa, Y. S. (2004). On the maximum drawdown of a Brownian motion. Journal of Applied Probability, 41(1), 147–161. https://doi.org/10.1239/jap/1077134610

Natixis Investment Managers. (2025). 2025 Wealth Industry Survey:The eve of disruption [White paper].Natixis Center for Investor Insight. https://www.im.natixis.com/content/dam/natixis/website/insights/investor-sentiment/2025/wealth-industry-survey/we alth-industry-survey-full-report.pdf

Philippon, T. (2016). The FinTech opportunity (NBER Working Paper 22476). National Bureau of Economic Research. https://doi.org/10.3386/w22476

Polygon.io. (2025). Stocks API v2: Aggregates endpoints documentation. https://polygon.io/docs/stocks/

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., & Dormann, N. (2021). Stable Baselines3. GitHub repository. https://github.com/DLR-RM/stable-baselines3

Sandberg, R. (2019). Momentum everywhere: Portfolios of global momentum strategies (S&P Global Quantamental Research White Paper).

Sharpe, W. F. (1966). Mutual fund performance. Journal of Business, 39(1), 119–138. https://doi.org/10.1086/294846

Sironi, P. (2016). FinTech innovation: From robo-advisors to goal-based investing and gamification. John Wiley & Sons.

Statista Research Department. (2023, April 15). Assets under management of robo-advisors worldwide from 2018 to 2024, with a forecast to 2028. Statista.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.

Xu, H., De Luca, S., & Rowley, J. (2024). Buying local: Measuring the persistence of home bias in U.S. equity portfolios (Vanguard Research Note). Vanguard.

Zhang, Y., & Aggarwal, C. C. (2024). Machine learning in portfolio management: A survey. Journal of Finance and Data Science, 7(3), 149–169.https://doi.org/10.1016/j.jfds.2023.08.001

Appendix

Data Provenance

- **Prices and volumes**. Split- and dividend-adjusted daily bars were downloaded from the polygon.io REST API on 14 June 2025.
- **Risk-free rate**. The 3-month U.S. Treasury bill secondary-market yield, obtained from the Federal Reserve's H.15 release.
- **Commission assumptions**. Zero explicit commission; round-trip slippage varied from 2–10 bps in sensitivity tests (see Section 4.4).

Asset Universe (01 Jan 2024 snapshot)

Ten Fractional-Eligible Securities Used in Back-tests

Ticker	Description	Sector / Theme
VOO	Vanguard S&P 500 ETF	Broad-market core
IJR	iShares Core S&P Small-Cap ETF	Small-cap core
ARKK	ARK Innovation ETF	Disruptive tech thematic
AAPL	Apple Inc.	Information Technology
MSFT	Microsoft Corp.	Information Technology
F	Ford Motor Co.	Consumer Discretionary
PFE	Pfizer Inc.	Health Care
PLTR	Palantir Technologies	Information Technology

Ticker	Description	Sector / Theme
BAC	Bank of America	Financials
TSLA	Tesla Inc.	Consumer Discretionary

Note: All tickers were priced below $50 on 2 Jan 2024 and supported fractional share trading at major U.S. brokers, satisfying the small-capital constraint.

Soft-Actor–Critic (SAC) Hyper-parameters

Parameter	Value	Rationale
Network architecture	(128, 128) MLP	Matches Heaton et al. (2017)
Activation	ReLU	Stable with price inputs
Replay buffer	100 000 steps	Covers ~1 year of data
Batch size	256	Empirical RL finance norm
Discount factor γ	0.99	Emphasizes long-run reward
Policy LR (Adam)	3×10^{-4}	Default SB3 value
Entropy α	Auto-tuned	Encourages exploration
Training episodes	1 000 000 steps	Convergence ΔSharpe < 1%
Observation window	60 trading days	Captures quarterly dynamics
Reward	One-day Sharpe	Balances return/volatility

Code & Reproducibility

1. All Python 3.12 source files `data_ingestion.py`, `features.py`, `train_rl.py`, `backtest.py`, `simulate_investors.py`, `stats_tests.py`, `sensitivity_analysis.py` are available at: https://github.com/ttvAKL/algorithmic-retail-portfolio

2. The trained SAC agent weights are archived as `sac_2024.zip` in the same repository.

3. A Docker image (`Dockerfile`) pins all dependencies (`pandas 1.5`, `numpy 1.26`, `stable-baselines3 2.2`, `TensorFlow 2.15`, etc.).

4. Running

```
$ docker build -t retail-rl.
$ docker run --rm retail-rl python backtest.py
```

reproduces every figure and CSV used in Sections 4–5.