Simulating an Alpha Rotation Trading System via Random Forest Ranking

Reporter: 葉致均

Report Date:12/24

MOTIVATION

Research Question:
Unlike conventional approaches focused on directional prediction (up/down), this study pivots to analyzing Relative Strength. Using Random Forest, we identify stocks that are outperforming the market. through a weekly rotation strategy of these top-tier assets, we aim to capture Alpha and beat the SPY benchmark in bull cycles.

Evolution of Methodology

Phase 1: Daily Prediction (日線級別預測)

- **Methodology:** Utilized **Bi-LSTM** to predict next-day price direction (Up/Down).
- **Result:** Accuracy stagnated at **51-52%**, which is statistically equivalent to random guessing.
- **Reason for Failure:**
  - **Random Walk Noise:** Daily data is plagued by "Random Walk" characteristics with an extremely low **Signal-to-Noise Ratio (SNR)**.
  - **Transaction Costs:** The high frequency of trading caused fees and slippage to **erode** all potential profits.

Evolution of Methodology

Phase 2: Volatility Targeting (波動率目標管理)

- **Methodology:** Implemented automatic **Risk Control** by deleveraging (reducing positions) when volatility spiked.
- **Result: Underperformed** the benchmark.
- **Reason for Failure:**
  - **Volatility Drag in Bull Markets:** In a strong bull run, rapid upward momentum is inherently accompanied by high volatility.
  - **Over-Hedging:** Excessive risk control caused the model to **prematurely exit positions** (selling too early), missing the bulk of the gains during NVDA's parabolic rally.

Evolution of Methodology

Phase 3 (Finalized): Weekly Relative Rotation (週線相對強弱輪動)

- **Revised Philosophy:**
  - **Timeframe Aggregation (降維打擊):** Shifted from Daily → **Weekly** data to effectively filter out high-frequency noise.
  - **Paradigm Shift (典範轉移):** Pivoted from predicting **Absolute Direction** (Price Levels) → predicting **Relative Ranking** (Strength).

Experimental Process

- **Data Acquisition (ETL):** Sourced data via Yahoo Finance API (The "Magnificent 7" Tech Giants + SPY + VIX).
- **Data Resampling:** Converted/Aggregated daily timeframe data into "Weekly" data.
- **Feature Engineering:** Constructed Relative Strength indicators and Excess Return labels.
- **Model Training:** Utilized Random Forest for binary classification and probability ranking.
- **Strategy Backtesting:** Simulated a "Top-K" momentum rotation strategy.

Key Technology & Code Implementation

Python

```python
# 關鍵特徵工程
# 1. 計算個股與大盤的週收益率
df_weekly['weekly_ret'] = df_weekly['close'].pct_change()
df_weekly['spy_ret'] = spy_weekly['spy_ret']

# 2. 定義 Alpha (相對強度)：個股收益 - 大盤收益
df_weekly['excess_ret'] = df_weekly['weekly_ret'] - df_weekly['spy_ret']

# 3. 定義預測目標 (Labeling)
# 如果下週的 Alpha > 0，標記為 1 (跑贏大盤)，否則為 0
df_weekly['target'] = (df_weekly['excess_ret'].shift(-1) > 0).astype(int)
```

Experimental Model

Model Architecture

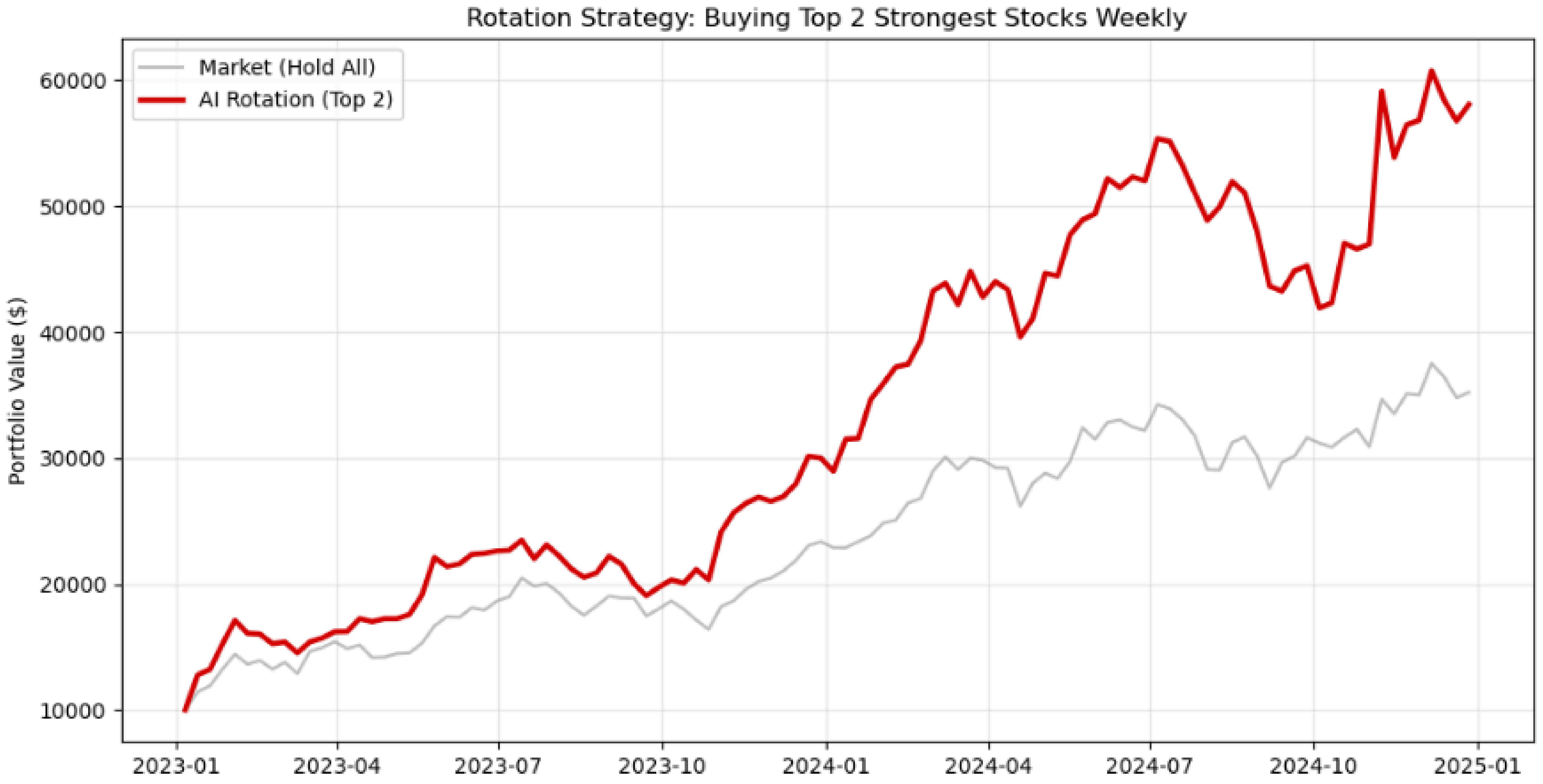
- **Model Selection:** Random Forest Classifier
- **Hyperparameters:**
  - `n_estimators=200`
  - `max_depth=5` (To prevent overfitting / For regularization)

DataSet

Dataset Overview

- **Time Period:** Jan 1, 2018 – Dec 31, 2024 (Encompassing the 2022 bear market and the 2023-24 bull run).
- **Frequency:** Weekly (Friday Close).
- **Assets (Tickers):** NVDA, AMD, TSLA, META, NFLX, COIN, MRNA.

Experimental Results



Experimental Results

principal: $10000 / 0.1% fee

```
1. 市場基準 (Hold All 7):
   - 最終資產: $35,214.48
   - 總獲利  : $25,214.48
   - 報酬率  : 252.14%
----------------------------------------
2. AI 輪動策略 (Top-2):
   - 最終資產: $58,080.92 🔥
   - 總獲利  : $48,080.92
   - 報酬率  : 480.81%
```

Results Analysis

Key Success Factors:

Portfolio Concentration: By investing exclusively in the Top-2 assets, we avoided the performance drag caused by mediocre stocks (the "average" performers).

Momentum Effect: The model successfully learned the principle that "strength begets strength." During NVDA's parabolic surge, the AI consistently assigned high confidence scores, allowing us to capture the bulk of the major rally.

Long-Only Approach: Adopting a strict Long-Only strategy proved crucial. In the powerful bull market of 2023-2024, this prevented the catastrophic losses often associated with shorting strong momentum stocks.

Feature Importance:

The model analysis reveals that **RSI** and **Momentum_4W** are the most critical predictors for forecasting next week's relative strength.

Significance of Key Parameters

- `TOP_K = 2` **(Weekly Position Count):** This defines the number of stocks held each week. Setting this to 2 strikes a balance between **risk diversification** and **profit concentration**.
- `resample('W-FRI')` **(Weekly Resampling):** Resampling data to a weekly frequency (Friday close). This was the single most critical parameter for performance, boosting the **predictive win rate** from ~51% on daily data to over **57%** on weekly data.
- `max_depth = 5` **(Decision Tree Depth):** Constraining the tree depth forces the model to learn **generalizable patterns** (general rules) rather than merely **memorizing** (overfitting) historical noise.

Comparison & Limitations

Comparative Analysis:

- **Our Approach:** Categorized as **Swing Trading**.
  - Operates at a **low frequency** (Weekly).
  - High **latency tolerance** (execution speed is not critical).
  - Relies primarily on **Closing Prices**.
- **Industry HFT/Quant:** Categorized as **High-Frequency Trading (HFT)** or **Intraday Trading**.
  - Operates at **ultra-high frequency** (Milliseconds/Microseconds).
  - Relies on **Order Book (Level 2 Data)**.
  - Extremely **latency-sensitive**.

Comparison & Limitations

Limitations:

- **Lack of Alternative Data:** The model currently relies solely on price-volume data, excluding **News Sentiment** or **Market Positioning** (Chip Analysis).
- **Simplified Transaction Cost Estimation:** While a 0.1% fee was factored in, the simulation does not account for **Slippage** or **Market Impact**.
- **Survivorship Bias:** The investment universe consists entirely of currently surviving tech giants, which may overstate historical performance.

Future Optimization

- **Integrate NLP Sentiment Analysis:** Scrape discussion volume from Reddit and Twitter to serve as a **leading indicator** of market sentiment and hype.
- **Risk Parity Allocation:** Shift from an equal-weight strategy to **Risk Parity**, assigning weights based on volatility (i.e., allocating less capital to highly volatile assets).
- **Enhance Macro Filters:** Incorporate Federal Reserve interest rate decisions or the Dollar Index (DXY) as a **"Master Switch"** (or Regime Filter) to control overall market exposure.

Summarize

- **Turning Noise into Alpha:** We leveraged Big Data to distill signal from the noise.
- **Choice > Effort:** Focusing on **Asset Ranking** yields better results than guessing **Absolute Price**.

Q&A

1. "Why is your backtest accuracy (Win Rate) only around 57%, yet the strategy generates such exceptionally high returns?"

2. "You selected only these 7 specific tech stocks. Doesn't this introduce Survivorship Bias? How would the strategy perform if the selection pool included stocks that were eventually delisted?"

3. "Why did you choose Random Forest instead of currently trending Deep Learning models like LSTM or Transformers?"

Conclusion

In the world of investing, running fast does not guarantee a win; surviving the longest is the key.

Thanks For Listening