Project Work – A Factor-based A Share Quantitative Trading Strategy

(Full code and output on github.com/joeylyz/py-factor-screening-a)

The project aim design and implement of a systematic China A share strategy, with focus on a mixed of both technical and fundamental factor screening, and perform back test. Then machine learnings tools are deloyed to explain the relationship between factors and real share price. In the final, we conducted a portfolio analysis with both risk and return parameters.

Part One: Universe Construction: get_china_stocks.py

How it works:

- Uses Tushare (a cleaned API with exchange information) to fetch all currently listed stocks in China.

- Applies filters on exchange, volume, market cap

Output:

- china_stocks.csv — the master list of all stocks to be considered for the strategy.

Part Two: Data Engineering & Strategy Backtest : china_stock_strategy.py

How it works:

- **Data Preparation**

  For each stock, fetches or loads Technical indicators (RSI, MACD, Bollinger Bands) and Fundamental indicators (PE, PB, Debt/Equity, ROE, Revenue Growth)

- **Scoring & Selection**

  At the beginning of each month, for each stock, calculates technical indicators (based on indicator thresholds) and fundamental indicators (based on value/quality thresholds), assigned a fixed weight to each score, and calculates total score.

  Then, ranks all stocks by score and selects the top 10 for the monthly portfolio.

- **Backtest Portfolio Construction**

  Simulates an equal-weighted portfolio of the top 10 stocks, rebalanced monthly. Calculates monthly returns and cumulative portfolio value.

Output:

- monthly_recommendations.csv – contains trade recommendations for each month

- Cumulative return of portfolio value vs. CSI300 (strategy_vs_csi300.png)

Part Three: Machine Learning Strategy: ml_strategy.py

How it works:

- **Data Preparation**

  Loads all monthly stock metrics and historical prices and returns. Prepares features (technical + fundamental) and target (next-month return).

- **Model Training**

  Trains a Random Forest regressor to predict next-month returns from features.

- **Model Evaluation**

  Evaluates model performance and feature importance.

- **Model Prediction and Portfolio Construction**

  For each month in the test period, used the trained model top redicts returns for all stocks using the trained model. Selects the top 10 predicted stocks for the portfolio. Simulates and plots the returns.

Output:

- monthly_recommendations.csv — top ML picks for each month.

- Cumulative return of portfolio value vs. CSI300 (ML strategy_vs_csi300.png)

- Trained model

Part Four: Performance Analysis: portfolio_analysis.py

How it works:

- **Performance Metrics Calculation**

  Calculates total return, annualized return, volatility, Sharpe ratio, max drawdown, win rate, average win/loss, skewness, kurtosis, and risk metrics (VaR, ES, Beta).

- **Visualization**

Output:

- Performance summary table, top 3 PnL contributors per month table

- plots for all key metrics.

Output Snapshot

Get universe

```
After removing Beijing stocks: 5146 stocks
After volume filtering: 444 stocks
After market value filtering: 260 stocks
Successfully saved 260 stocks to china_stocks.csv
```

Factor model strategy back test performance vs CSI300 (2023 - 2024)
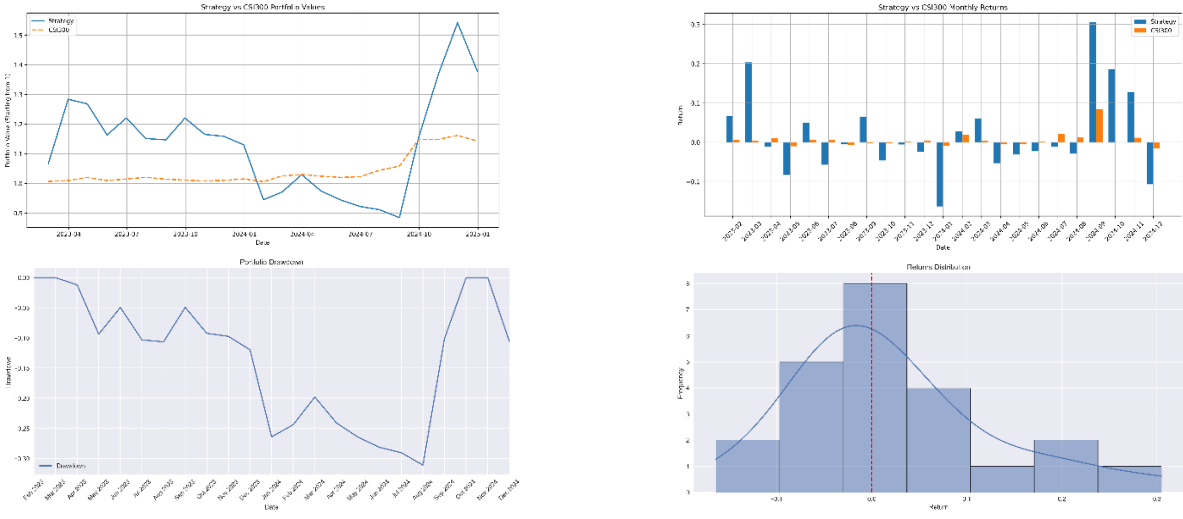

```
Strategy Performance:
Total Return: 37.83%
Annualized Return: 18.22%
Average Monthly Return: 1.91%
Monthly Return Std Dev: 10.39%
```

```
Performance Comparison:
Strategy Total Return: 37.83%
CSI300 Total Return: 14.30%
Strategy Annualized Return: 18.22%
CSI300 Annualized Return: 7.22%
Strategy Sharpe Ratio: 0.64
CSI300 Sharpe Ratio: 1.09
```

```
Monthly Trading Table:
------------------------------------------------------------
Date       Stock Code   Stock Name        Score
------------------------------------------------------------
2023-01    002261.SZ    拓维信息           2.36
2023-01    002031.SZ    巨轮智能           1.85
2023-01    300773.SH    拉卡拉           1.85
2023-01    002640.SZ    跨境通           1.85
2023-01    600518.SH    康美药业           1.70
2023-01    300225.SZ    金力泰           1.70
2023-01    600050.SH    中国联通           1.68
2023-01    002276.SZ    万马股份           1.68
2023-01    002027.SZ    分众传媒           1.53
2023-01    600828.SH    茂业商业           1.52
------------------------------------------------------------
2023-02    002261.SZ    拓维信息           2.36
```

Return and Risk Metrics

```
Return Metrics:
Total Return: 37.83%
Annualized Return: 18.22%
Annualized Volatility: 36.79%
Sharpe Ratio: 0.62
Win Rate: 39.13%
Average Win: 12.09%
Average Loss: -4.64%
```

```
Risk Metrics:
Maximum Drawdown: -31.10%
VaR (95%): -10.36%
VaR (99%): -15.12%
Expected Shortfall (95%): -13.50%
Expected Shortfall (99%): -16.40%
Beta: -3.75
```

```
Distribution Statistics:
Skewness: 0.97
Kurtosis: 0.85
```

Machine Learning model strategy performance vs CSI 300 (2025.01 ~)


```
Feature Importance:
            Feature   Importance
2       bb_position   0.185201
1         macd_hist   0.177698
0               rsi   0.136157
3          pe_ratio   0.124426
4          pb_ratio   0.114490
6               roe   0.092901
7    revenue_growth   0.086402
```

```
Model Evaluation:
Mean Squared Error: 0.0258
R-squared Score: -0.0392
```

```
Recommendations for 2025-01:
--------------------------------------------------------------
Stock Code   Stock Name      Predicted Return
--------------------------------------------------------------
002131.SZ    利欧股份           65.57%
300059.SZ    东方财富           54.90%
601162.SH    天风证券           52.01%
002640.SZ    跨境通           50.84%
000564.SZ    供销大集           49.47%
300180.SZ    华峰超纤           42.16%
600186.SH    莲花控股           41.39%
601933.SH    永辉超市           38.90%
600490.SH    鹏欣资源           38.28%
002031.SZ    巨轮智能           37.03%
```

```
Performance Metrics (Jan-Apr 2025):
Strategy Total Return: -4.65%
CSI300 Total Return: -0.21%
Strategy Sharpe Ratio: 1.50
CSI300 Sharpe Ratio: 3.30
```