8-Factors Model by Machine Learning

Ma Yadong

2017/4/17

Intraduction

Predict alpha returns with Machine Learning

- Traditional factor models are useful in portfolio selection
- Predicting alpha returns are of importance in factor models.
- Machine learning may give a better solution for prediction.
- LinearRegression, LogisticRegression, RandomForest_Classfier, RandomForest_Regression, Adaboost_Classfier, Adaboost_Regression

What are factor models

- CAPM

$$E(r_i) = \beta_0 + \beta_1 r_m + \varepsilon_i$$

- Fama-French 3 Factors

$$r_i = \beta_0 + \beta_1 r_m + \beta_2 r_{HML} + \beta_3 r_{SMB} + \varepsilon_i$$

- Multi-factor model

$$r_i = \sum \beta_i f_i + \varepsilon_i$$

Datasets

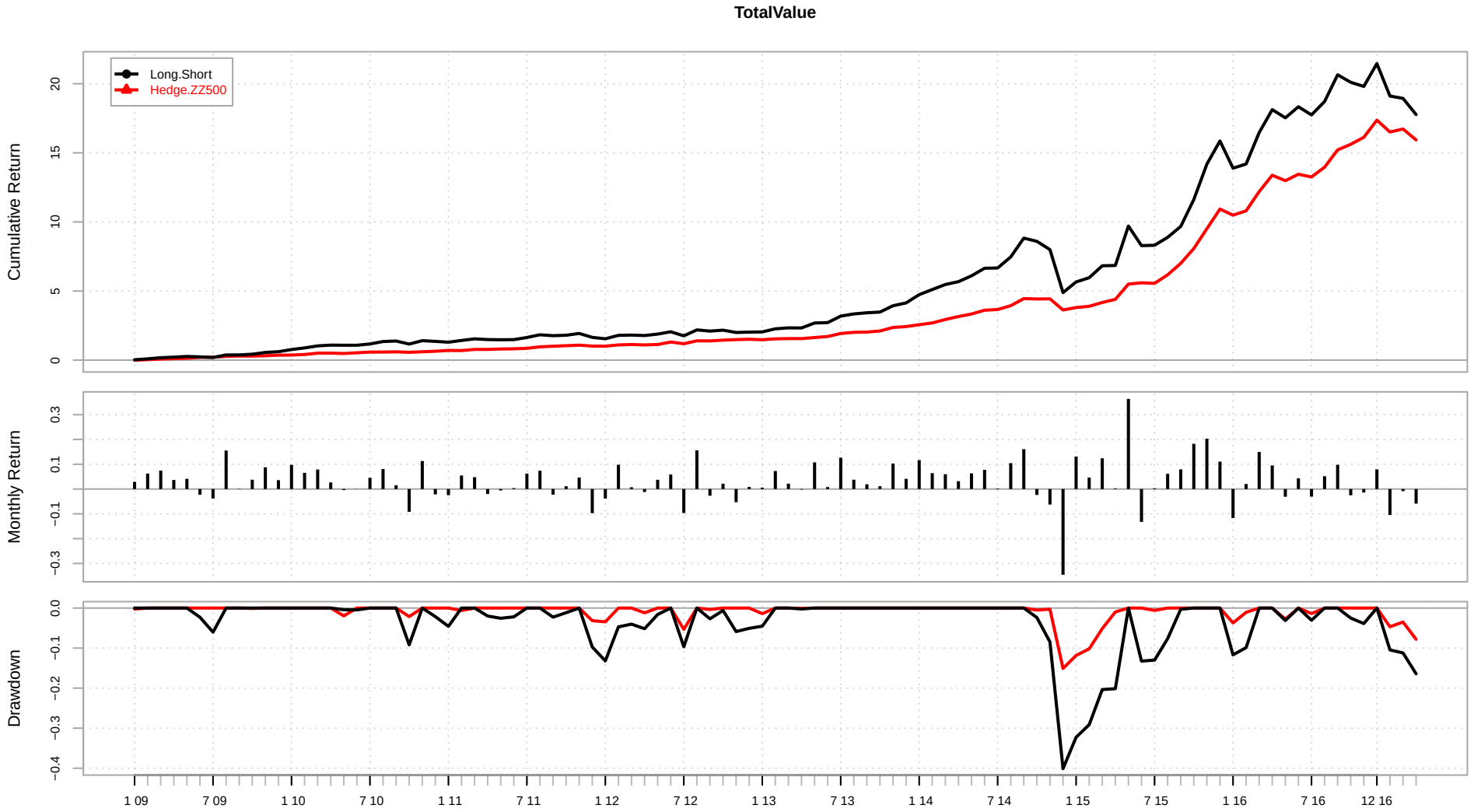
- The dataset are **daily data** in China stock market from **2009-01-01** to **2017-03-31**
- Trading strategy:
  1. At each trade times, select the best K stocks from the whole market according to **Scores**
  2. ST stocks and price limit stocks are not included in our targets
  3. Change the position monthly (20 trade days in fact)
  4. The weight of each stocks aresimplified as **equivalent weight**.
  5. At each trading, we use **the past 12 months** data to train our model and then to predict the scores.
- Scores:
  1. We are going to divide the stocks into **GOOD** and **BAD** two groups (with labels 1 and 0 correspondingly)
  2. We rank the return of future 20 days. We label the first 30% as 1 (GOOD), and last 30% as 0 (BAD)
  3. After the each prediction of our claffier, **the predicted probablity of fall into GOOD group will be the scores**.

Factors

1. **TotalValue**: the total capital value of stocks
2. **Reversal_DongFang**: a revesal indicator offered by DongFang Security
3. **Seasonal_NetProfitGrowth_YOY**: Net profit growth rate seasonally
4. **Turnover_ols_TotalValue**: value neutral turnover rate
5. **PB_lf**: latest Price bookvalue ratio
6. **HeteroRsquare_FF_d20**: Fama-French regression R^2
7. **PreRet_M_IndusRet_Citic1_m1**: Industrial neutral 1M monmentum
8. **PreRet_M_IndusRet_Citic1_m3**: Industrial neutral 3M monmentum

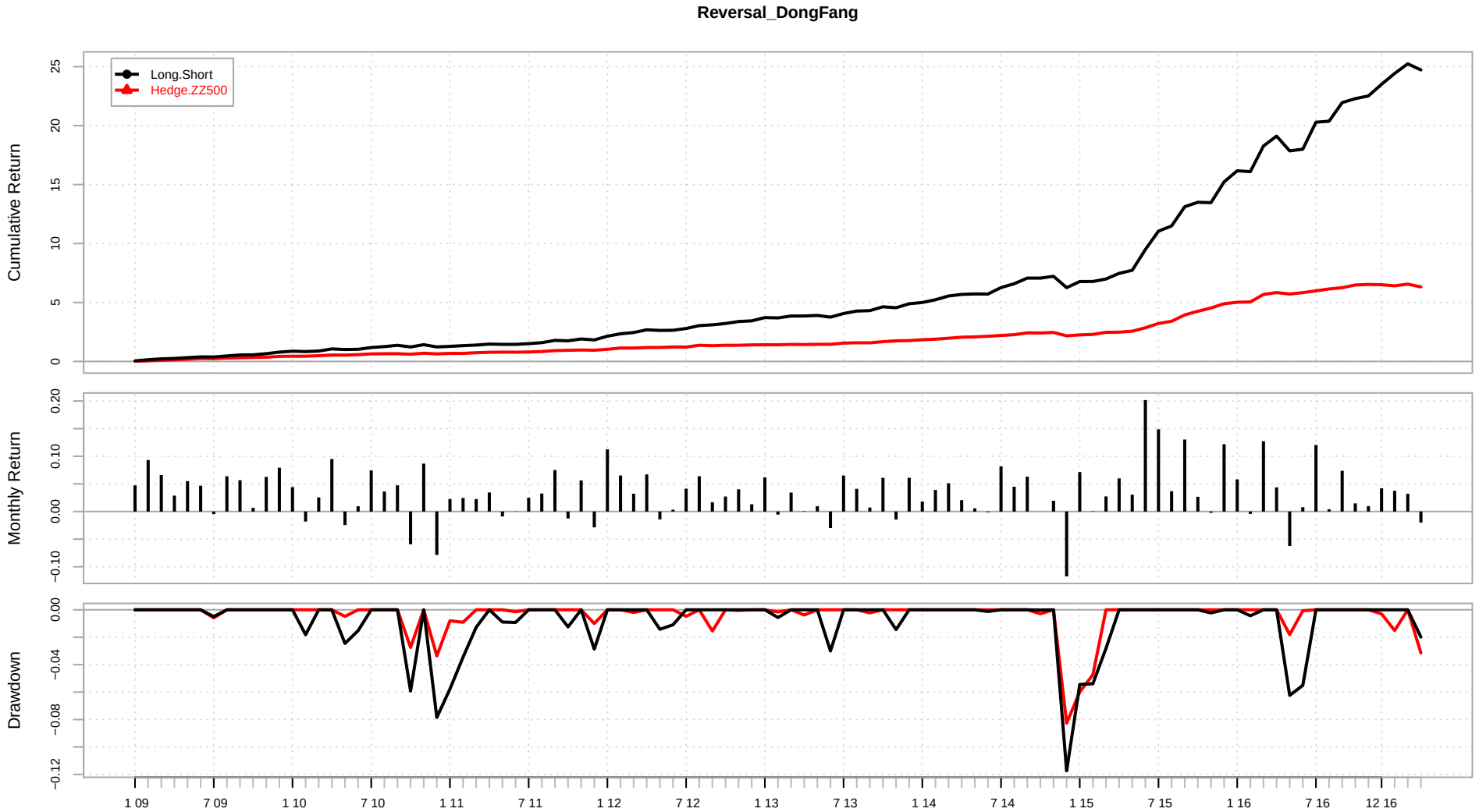
Single Factor Test Overview

TotalValue



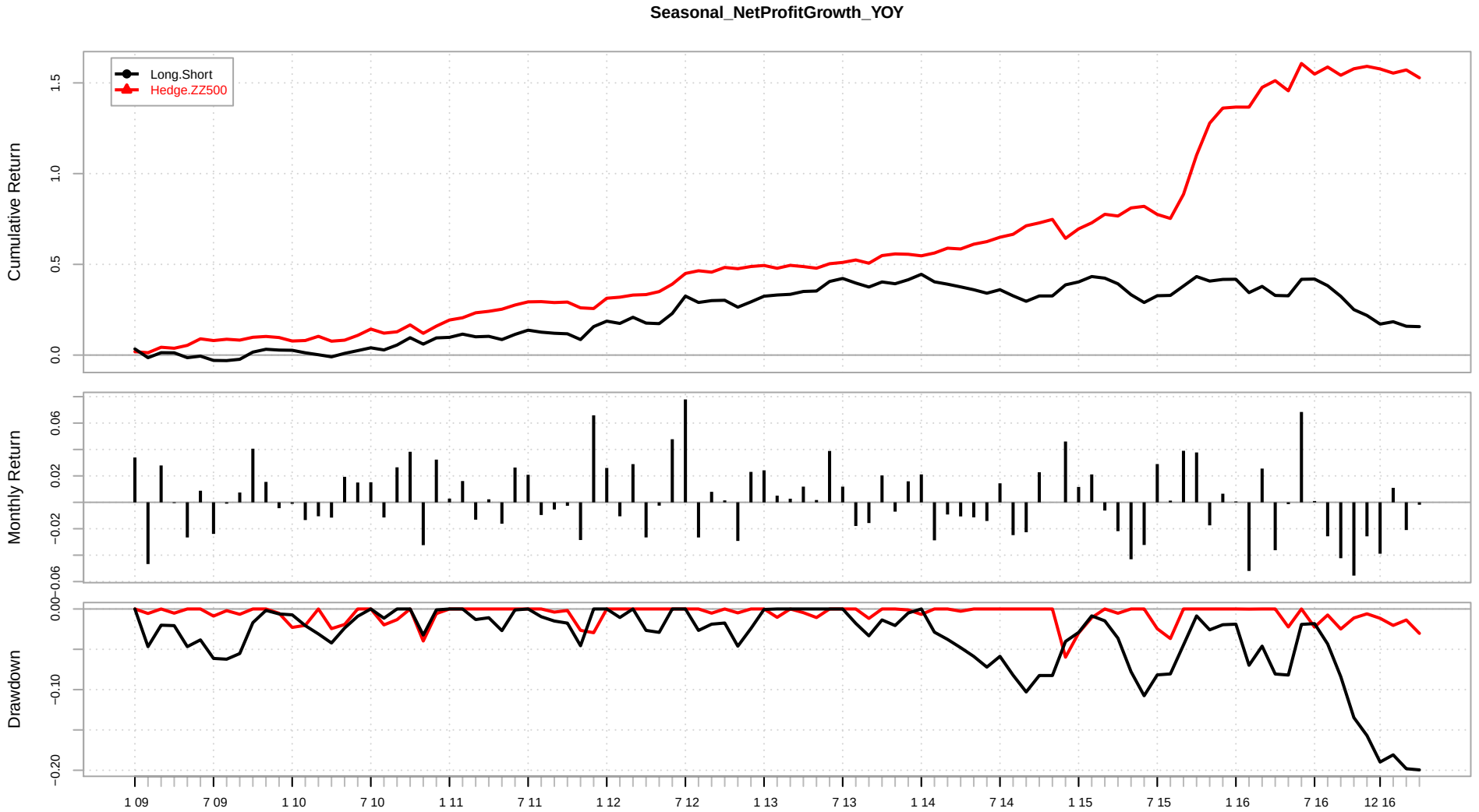
- Ret:40.91%,MaxDrawdown:15.1%,Sharpe ratio:2.218

Reversal_DongFang



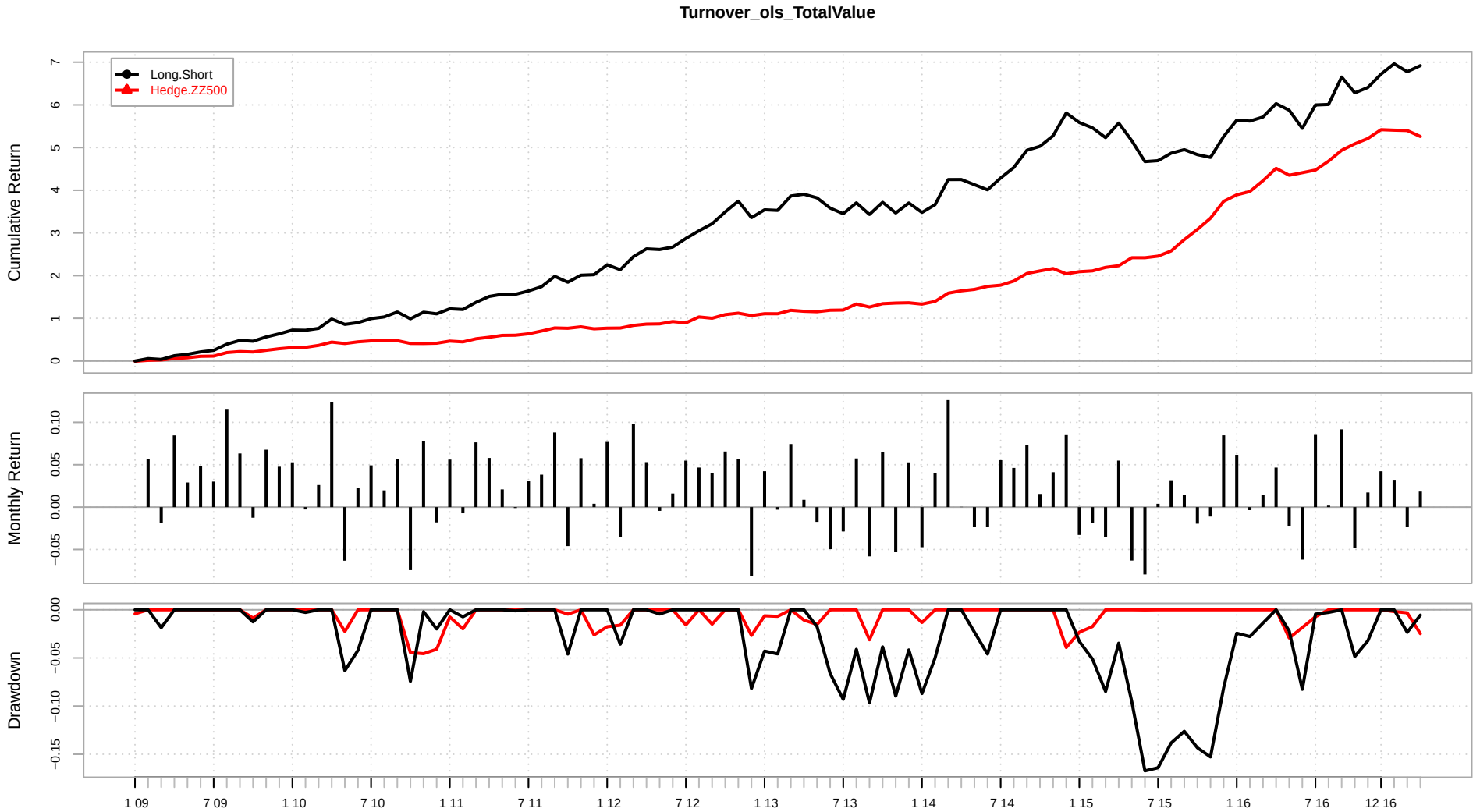
- Ret:27.29%,MaxDrawdown:8.26%,Sharpe ratio:2.517

Seasonal_NetProfitGrowth_YOY



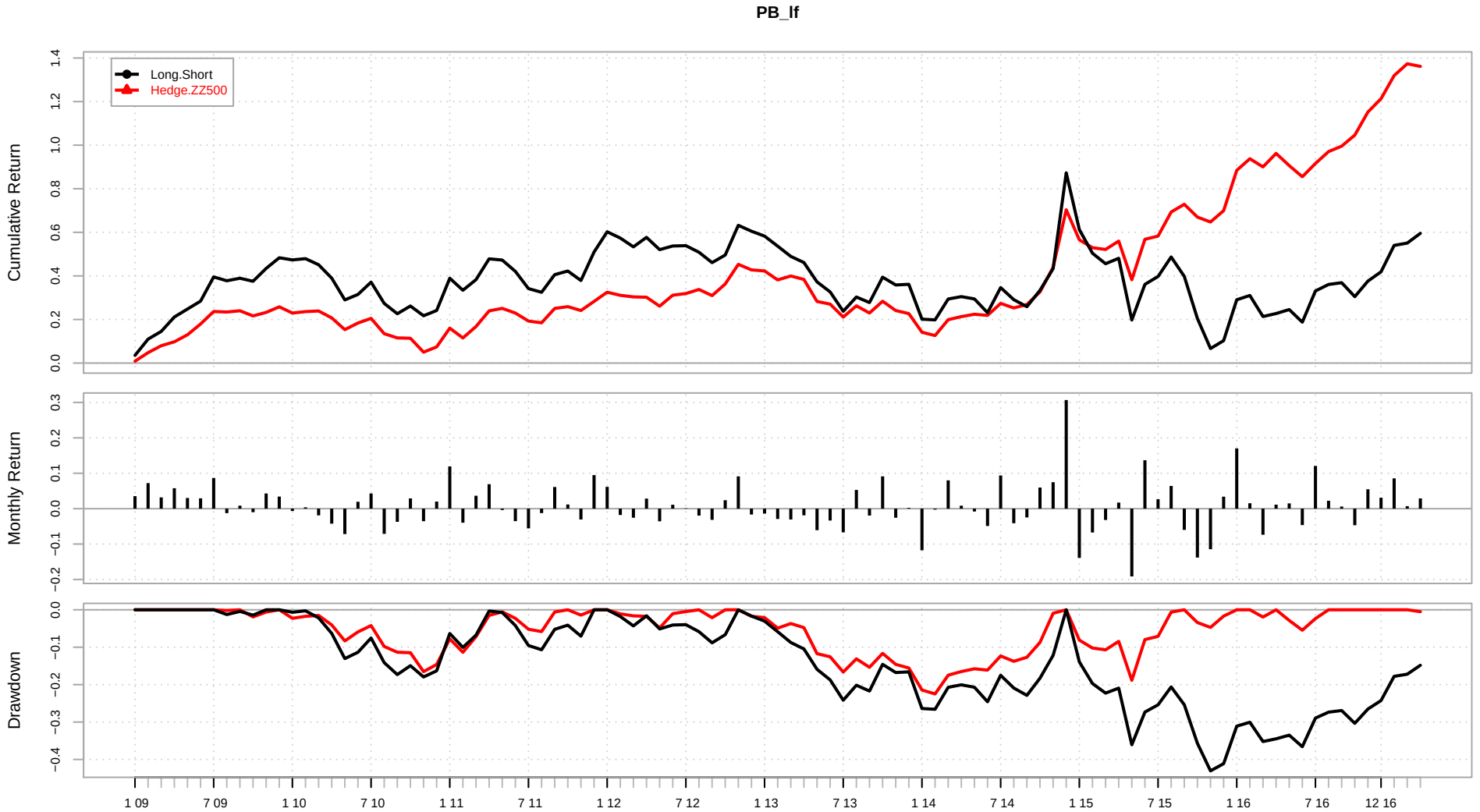
- Ret:11.9%,MaxDrawdown:5.98%,Sharpe ratio:1.404

Turnover_ols_TotalValue



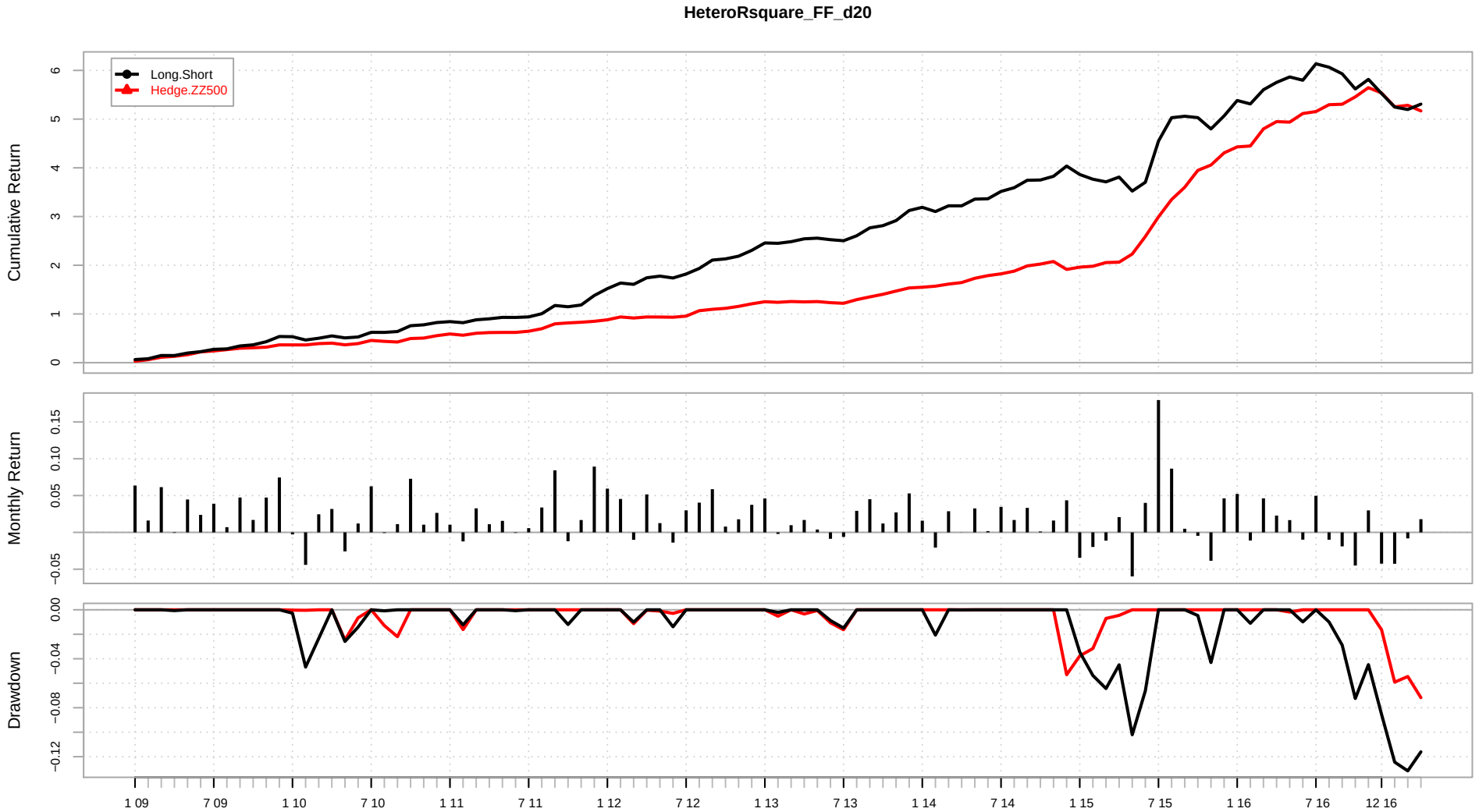
- Ret:24.9%,MaxDrawdown:4.55%,Sharpe ratio:2.45

PB_lf



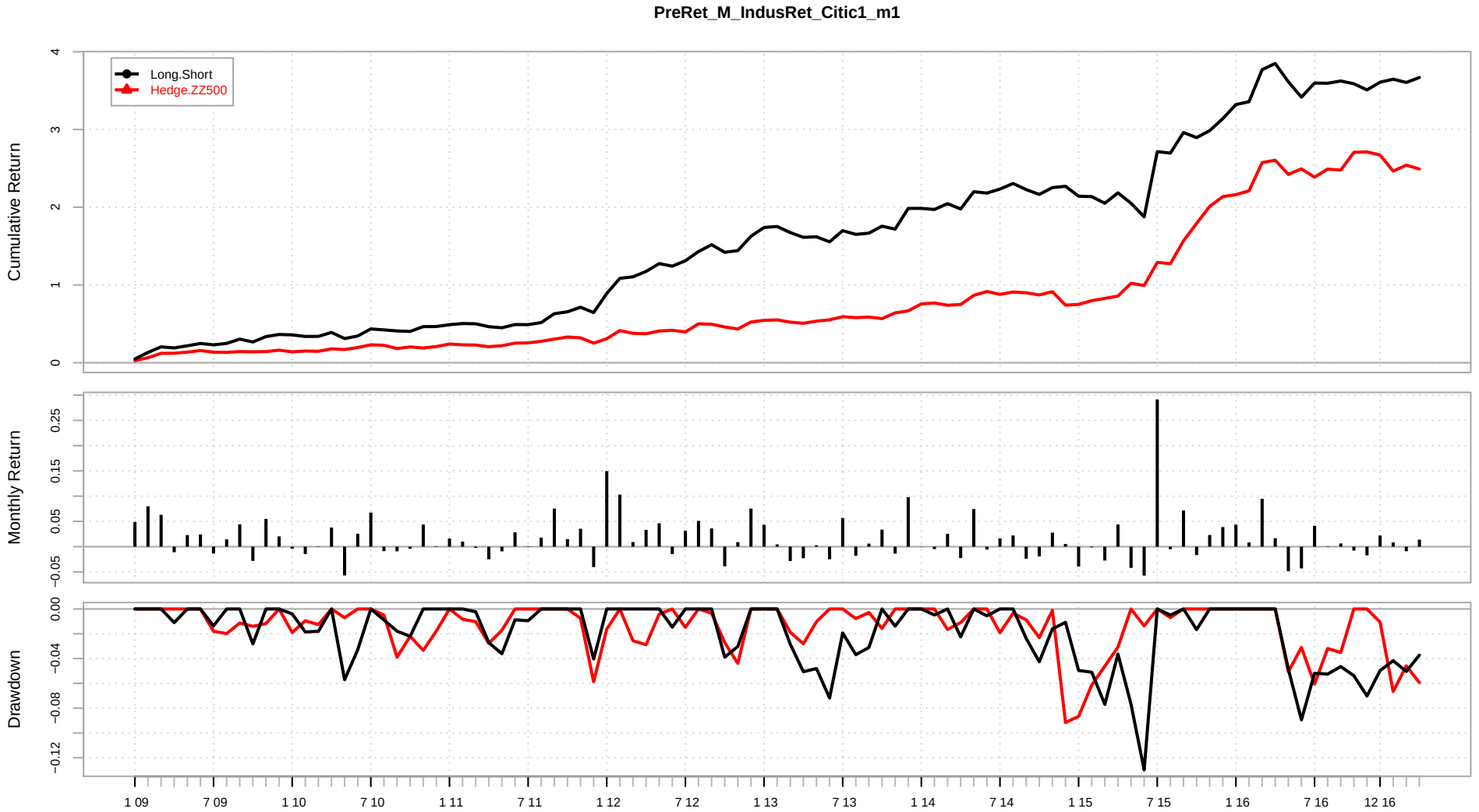
- Ret:10.98%,MaxDrawdown:22.49%,Sharpe ratio:0.772

HeteroRsquare_FF_d20



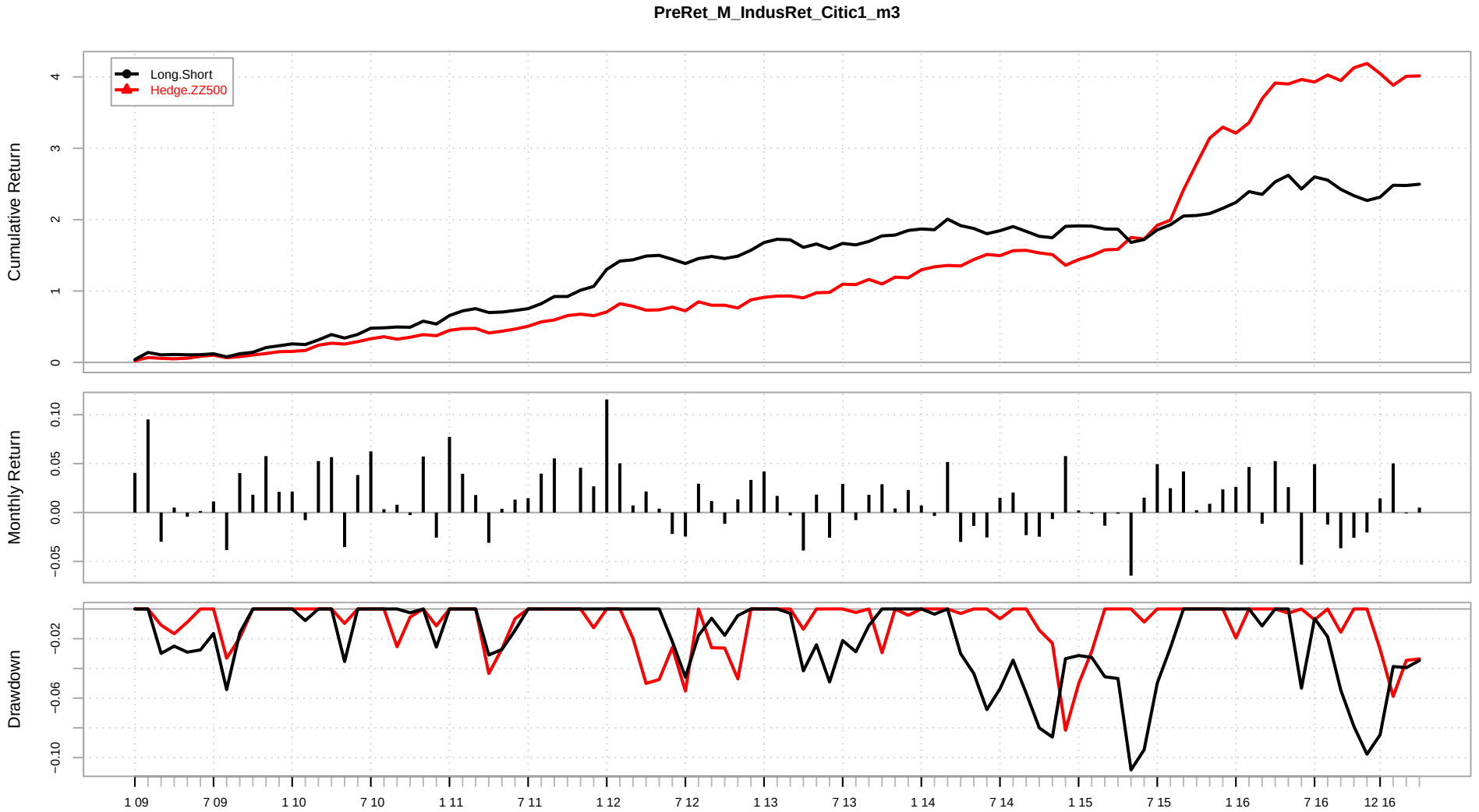
- Ret:24.67%,MaxDrawdown:7.18%,Sharpe ratio:2.514

PreRet_M_IndusRet_Citic1_m1



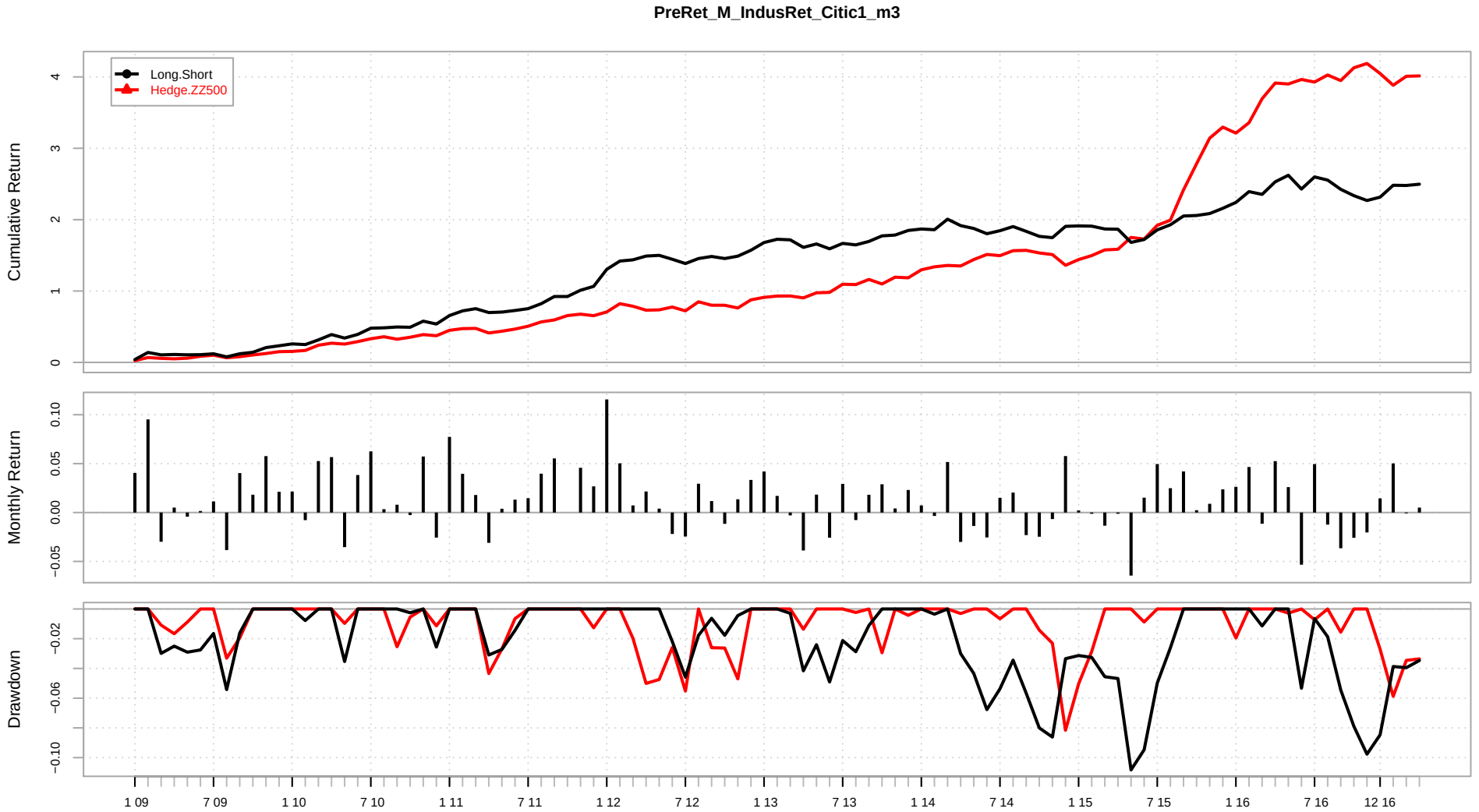
- Ret:16.36%,MaxDrawdown:9.16%,Sharpe ratio:1.262

PreRet_M_IndusRet_Citic1_m3



- Ret:21.58%,MaxDrawdown:8.16%,Sharpe ratio:1.795

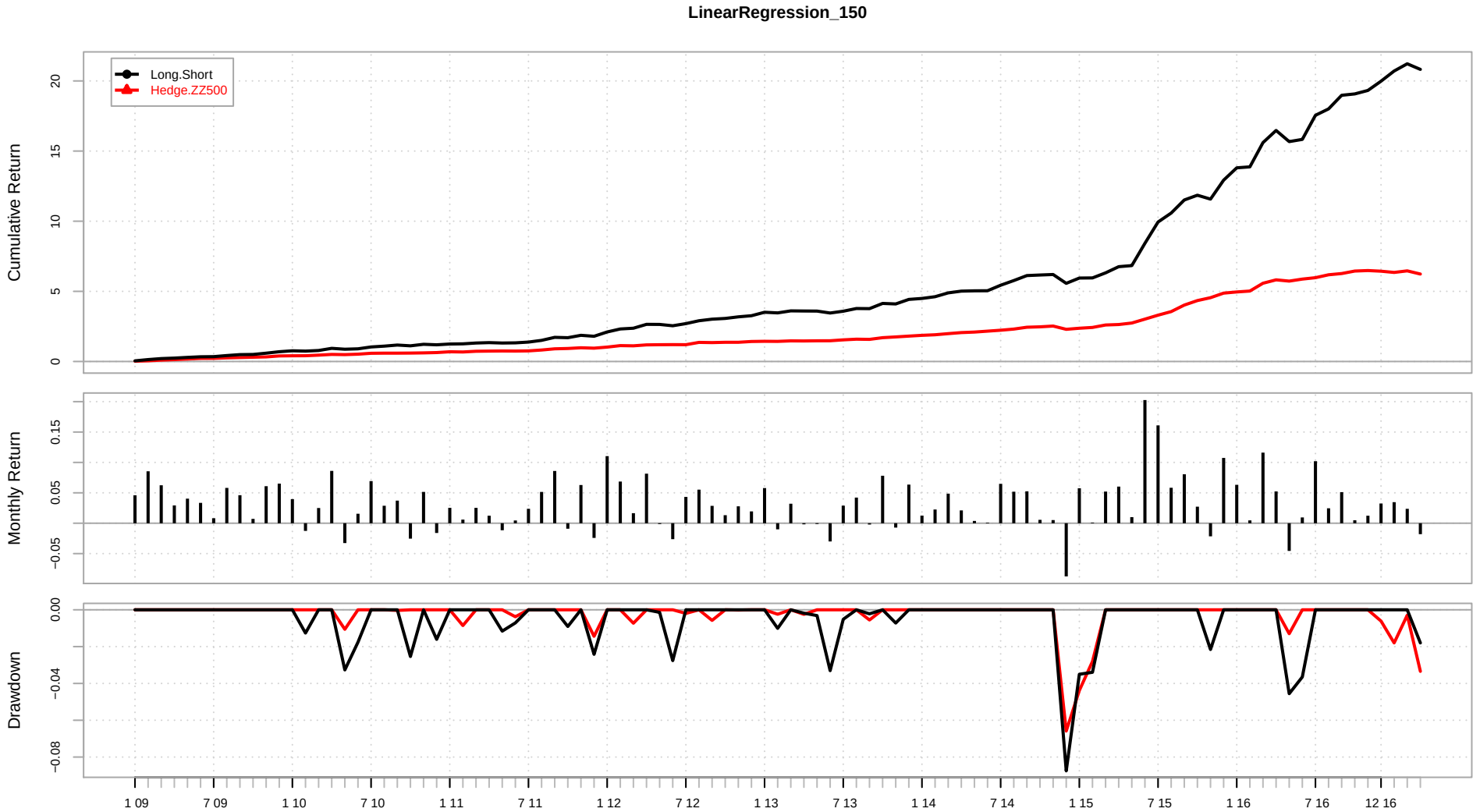
PreRet_M_IndusRet_Citic1_m3



- Ret:21.58%,MaxDrawdown:8.16%,Sharpe ratio:1.795

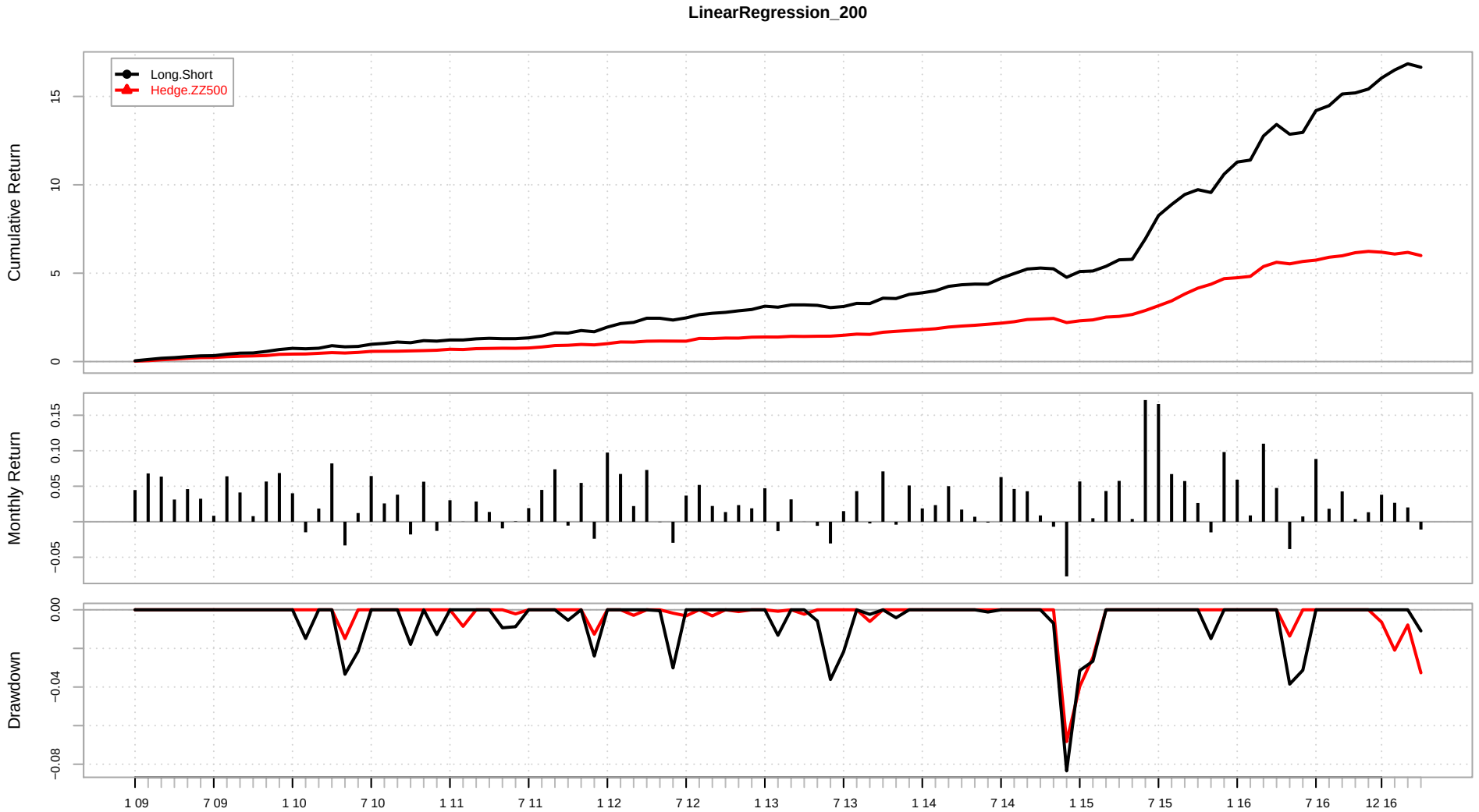
Machine Learning methods results

LinearRegression_150_EW



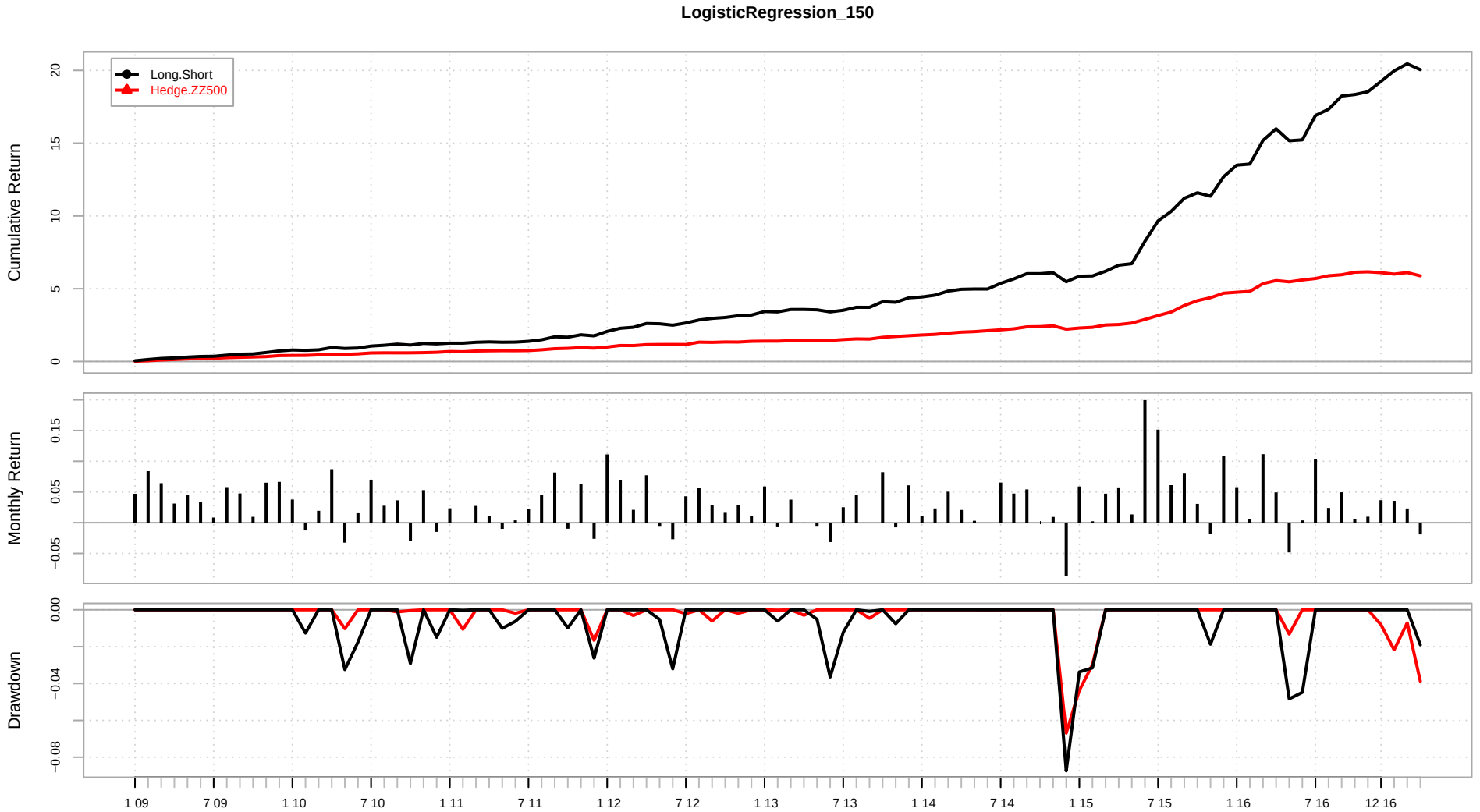
- Ret:27.1%,MaxDrawdown:6.59%,Sharpe ratio:2.875,Score:55.98%.

LinearRegression_200_EW



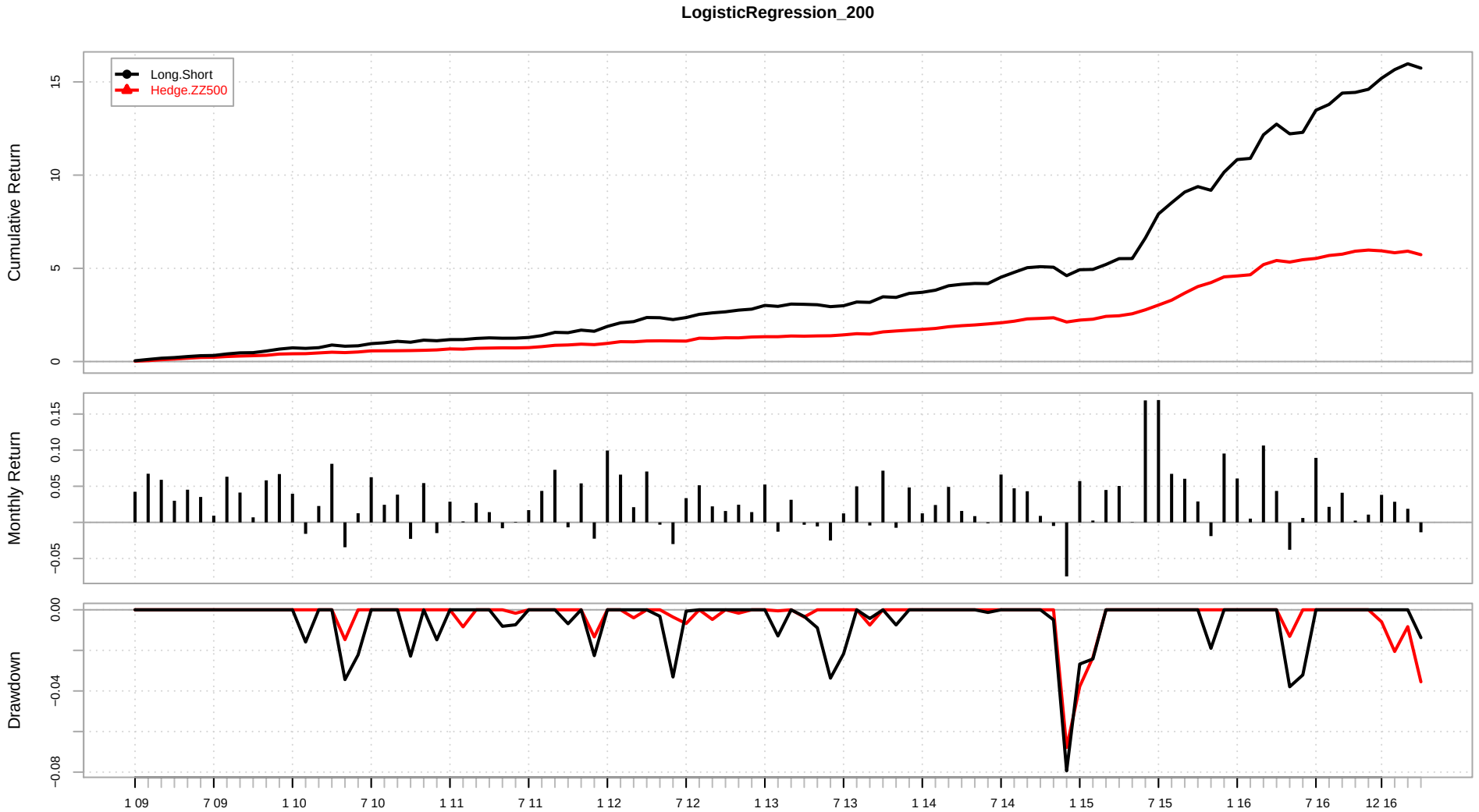
- Ret:26.6%,MaxDrawdown:6.83%,Sharpe ratio:2.918,Score:55.98%.

LogisticRegression_150_EW



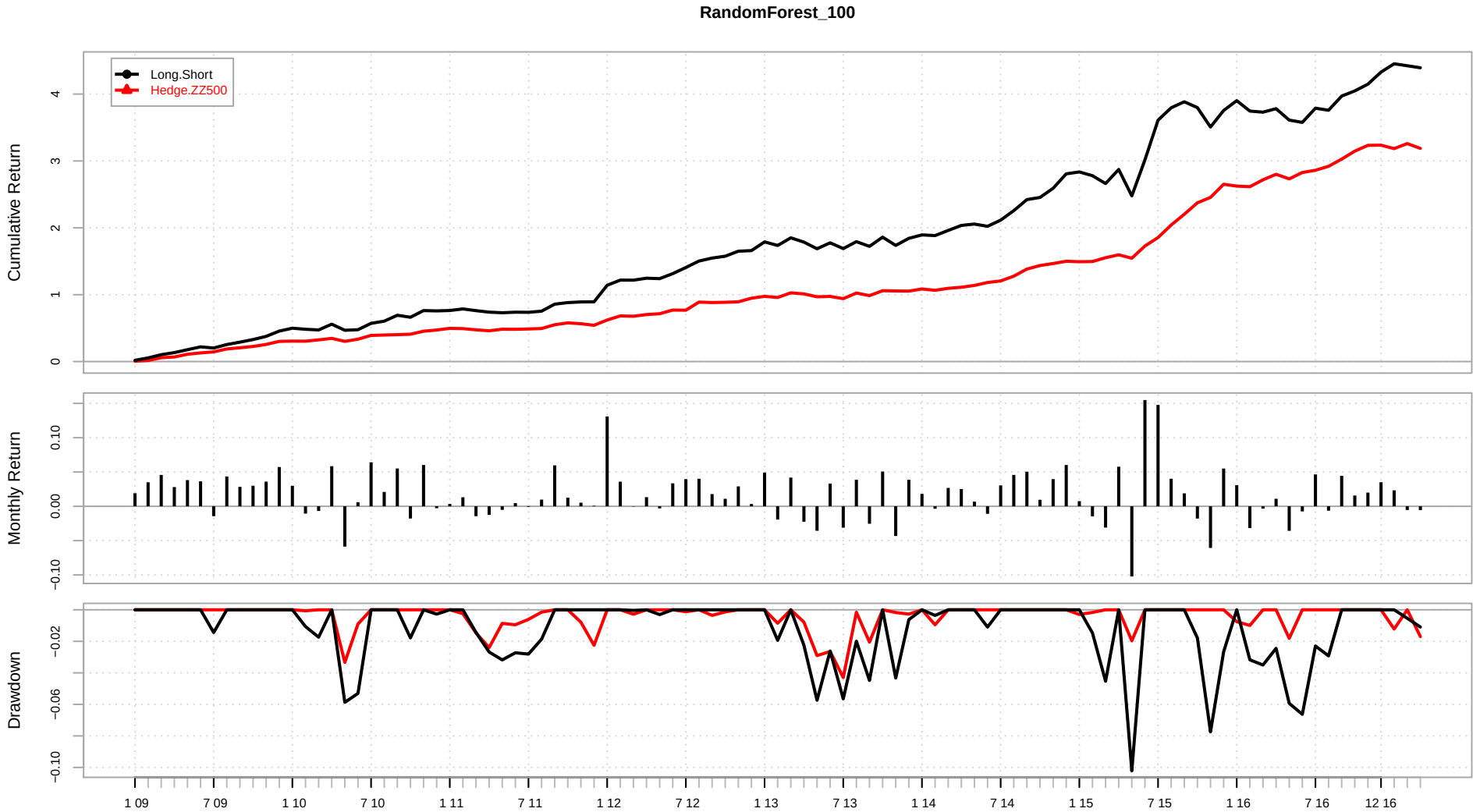
- Ret:26.34%,MaxDrawdown:6.69%,Sharpe ratio:2.809,Score:56.02%.

LogisticRegression_200_EW



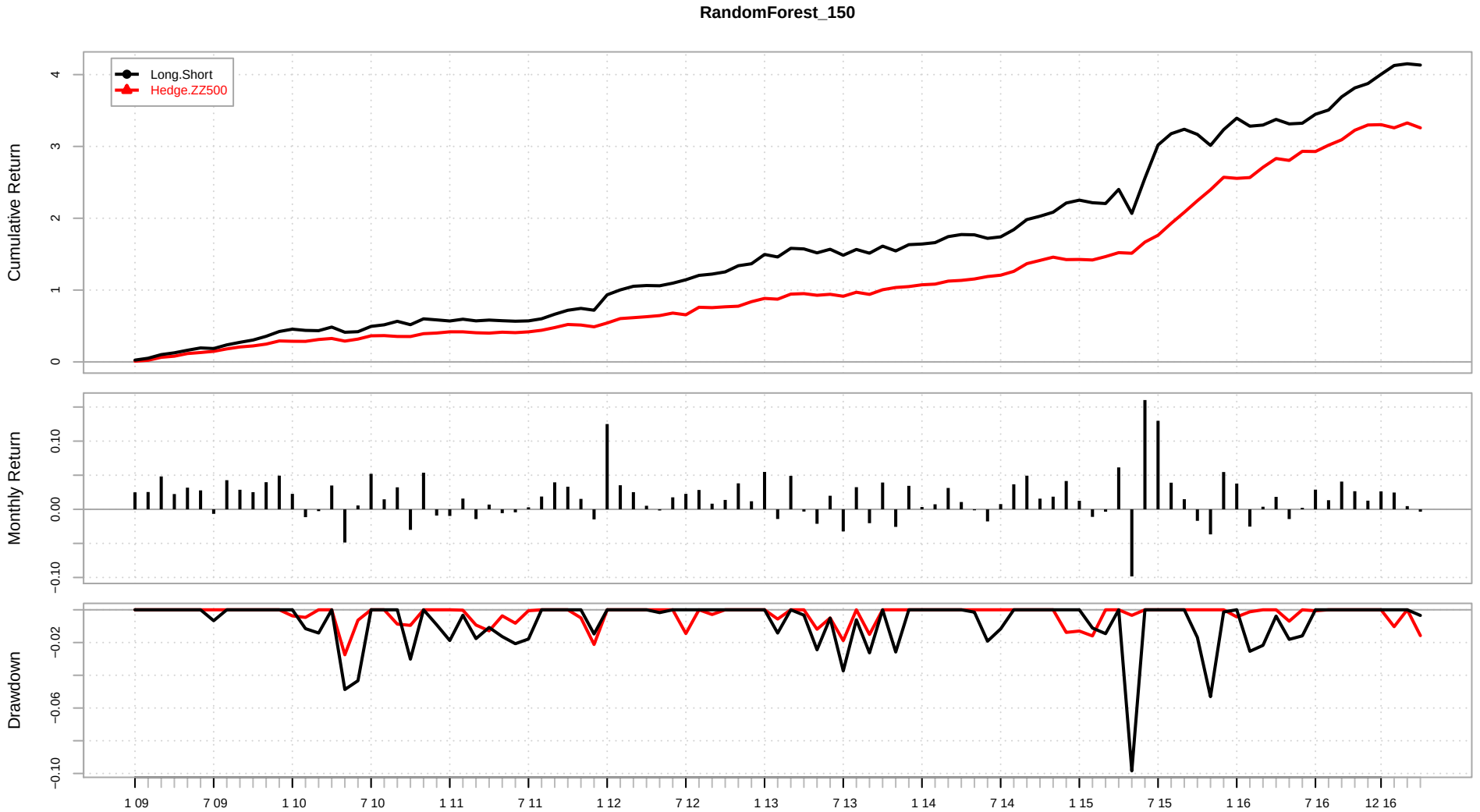
- Ret:26%,MaxDrawdown:6.79%,Sharpe ratio:2.843,Score:56.02%.

RandomForest_100_EW



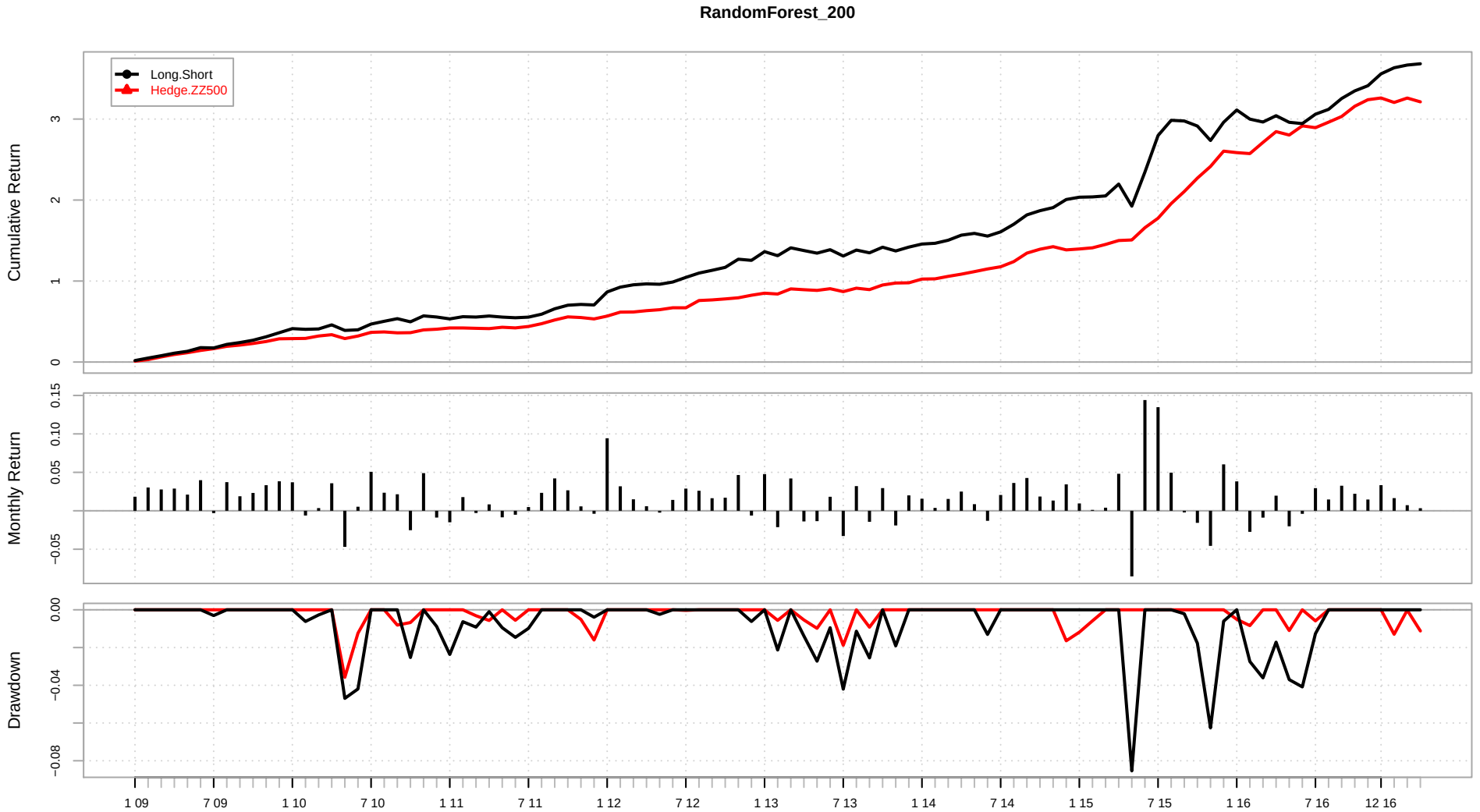
- Ret:18.96%,MaxDrawdown:4.3%,Sharpe ratio:2.415,Score:53.1%.

RandomForest_150_EW



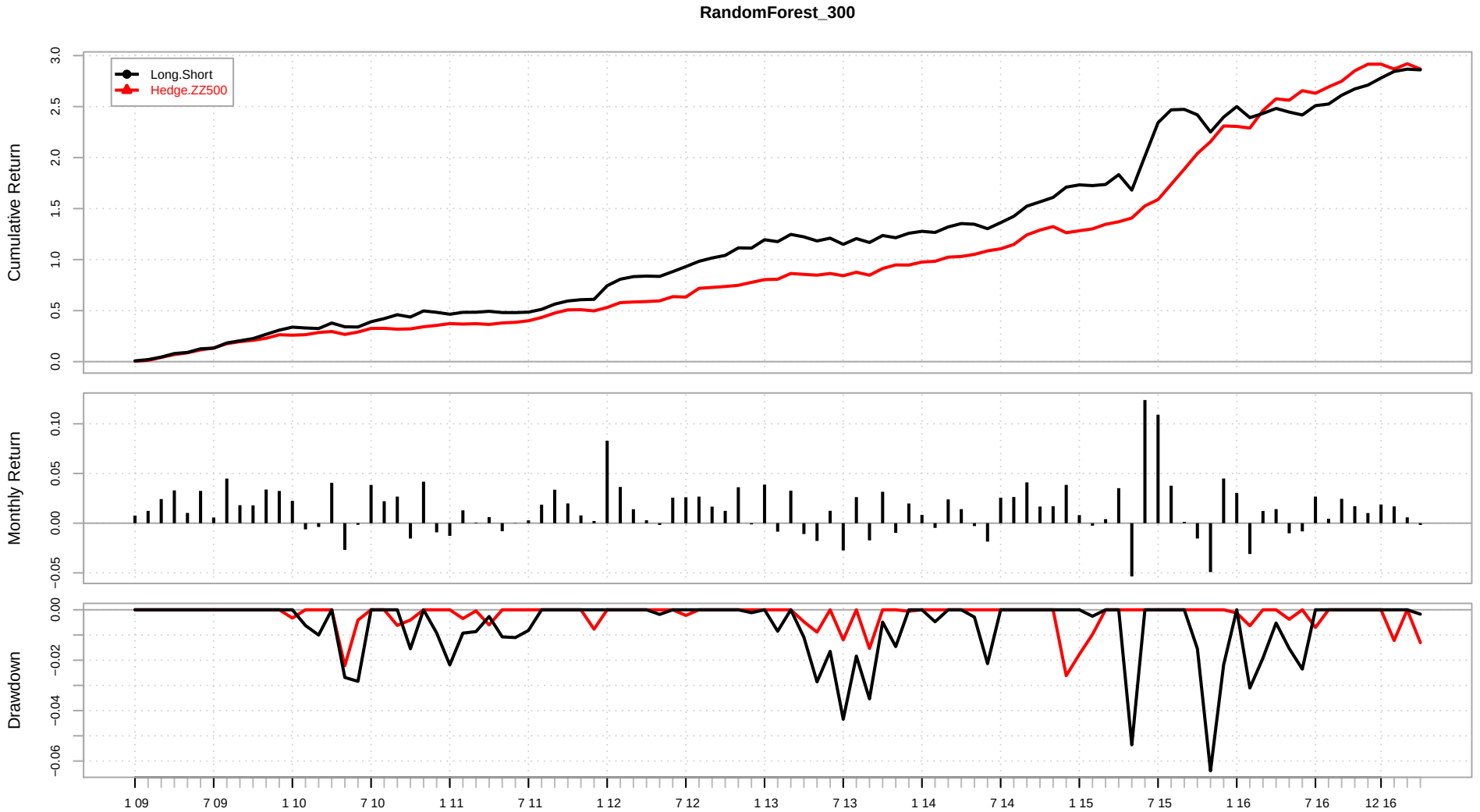
- Ret:19.2%,MaxDrawdown:2.75%,Sharpe ratio:2.676,Score:53.1%.

RandomForest_200_EW



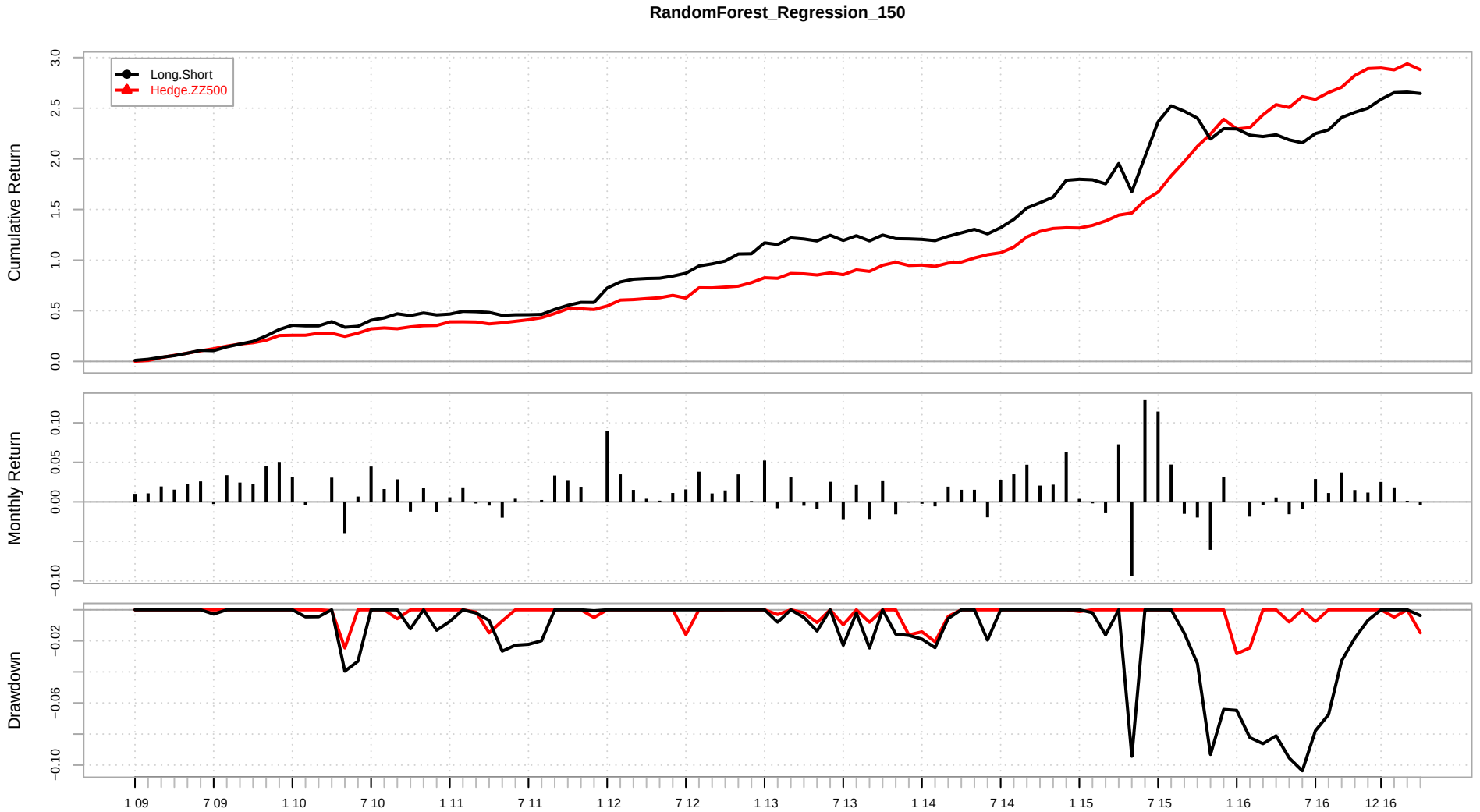
- Ret:19.04%,MaxDrawdown:3.58%,Sharpe ratio:2.792,Score:53.1%.

RandomForest_300_EW



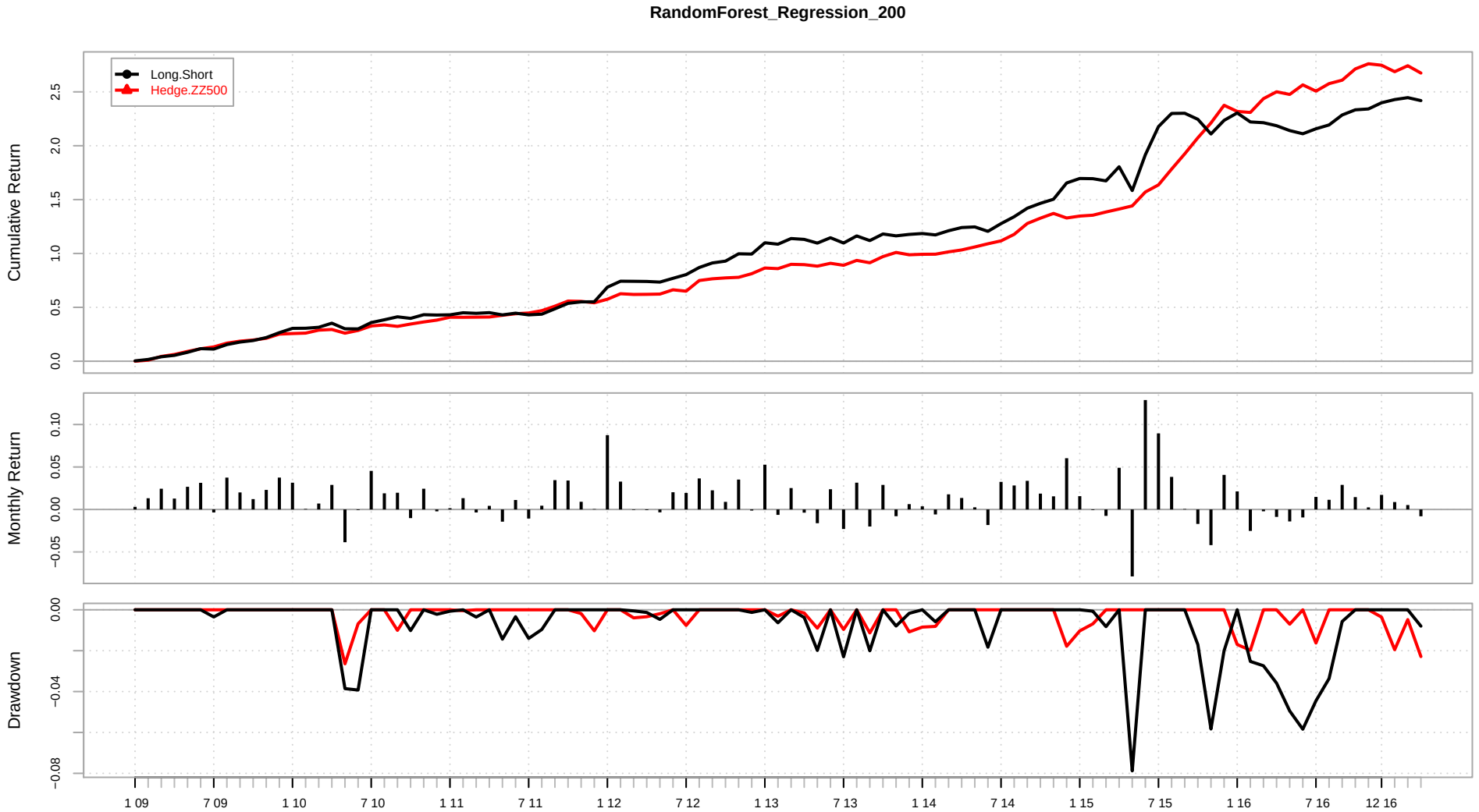
- Ret:17.82%,MaxDrawdown:2.61%,Sharpe ratio:2.819,Score:53.1%.

RandomForest_Regression_150_EW



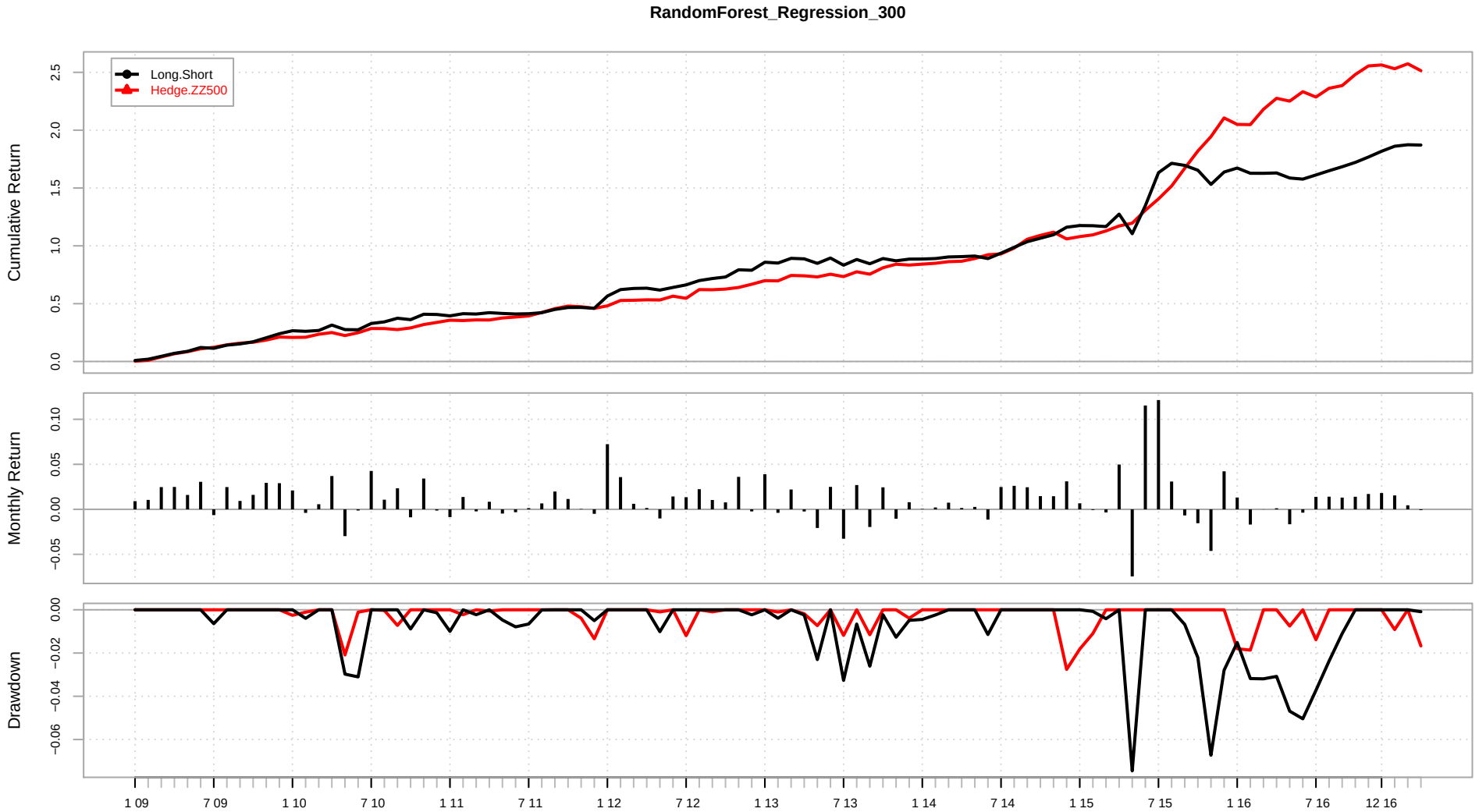
- Ret:17.86%,MaxDrawdown:2.83%,Sharpe ratio:2.698,Score:52.65%.

RandomForest_Regression_200_EW



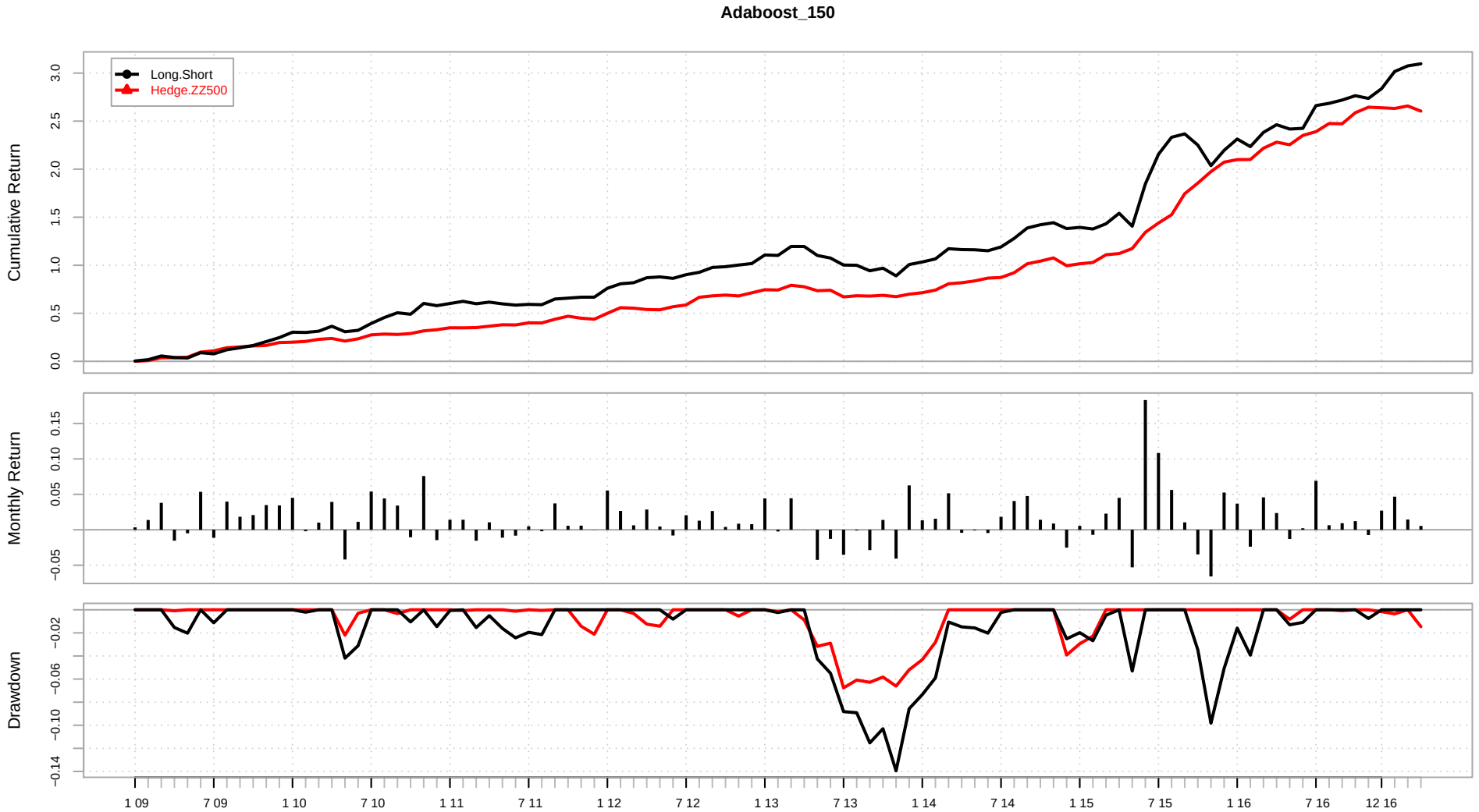
- Ret:17.09%,MaxDrawdown:2.64%,Sharpe ratio:2.603,Score:52.65%.

RandomForest_Regression_300_EW



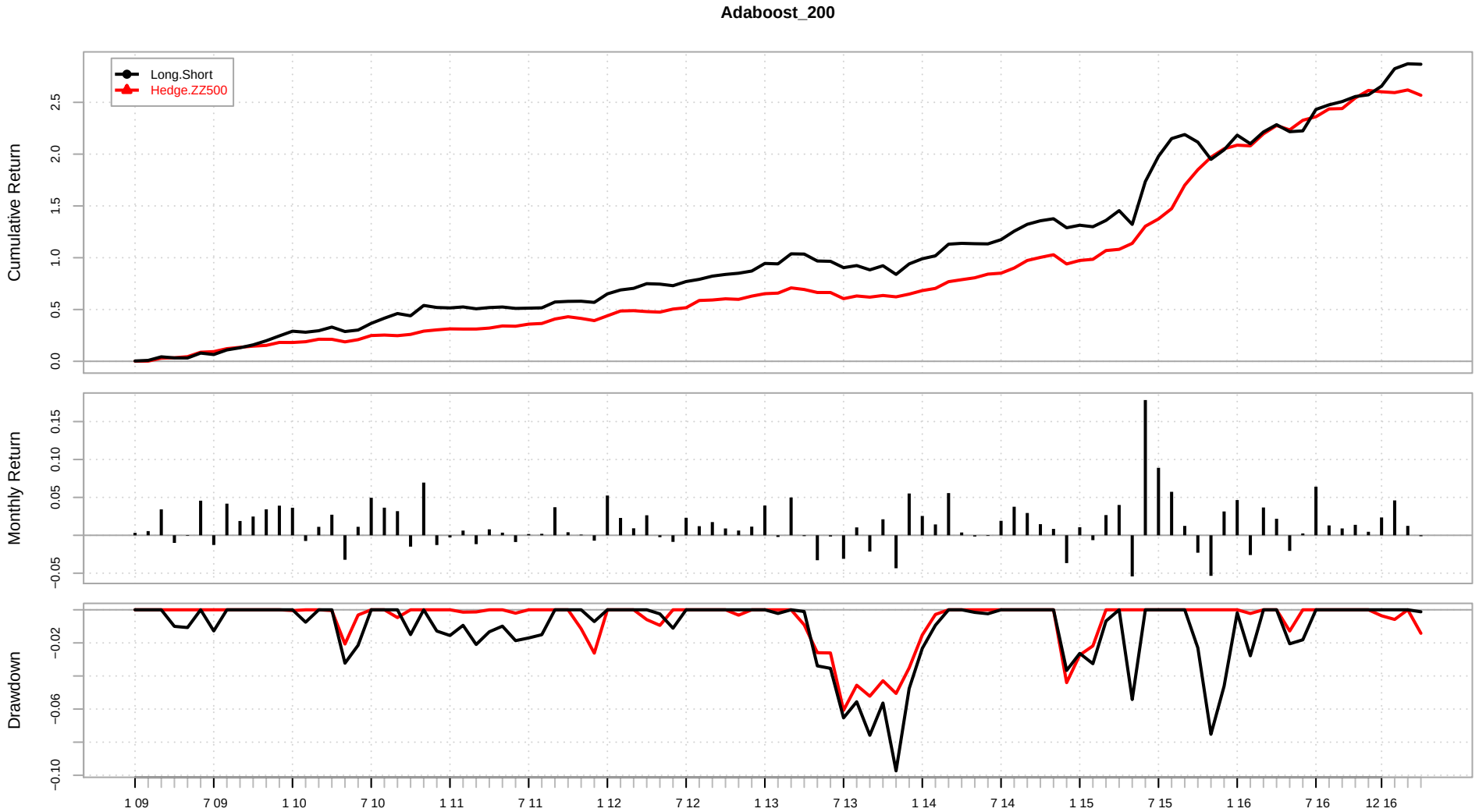
- Ret:16.46%,MaxDrawdown:2.76%,Sharpe ratio:2.563,Score:52.65%.

Adaboost_150_EW



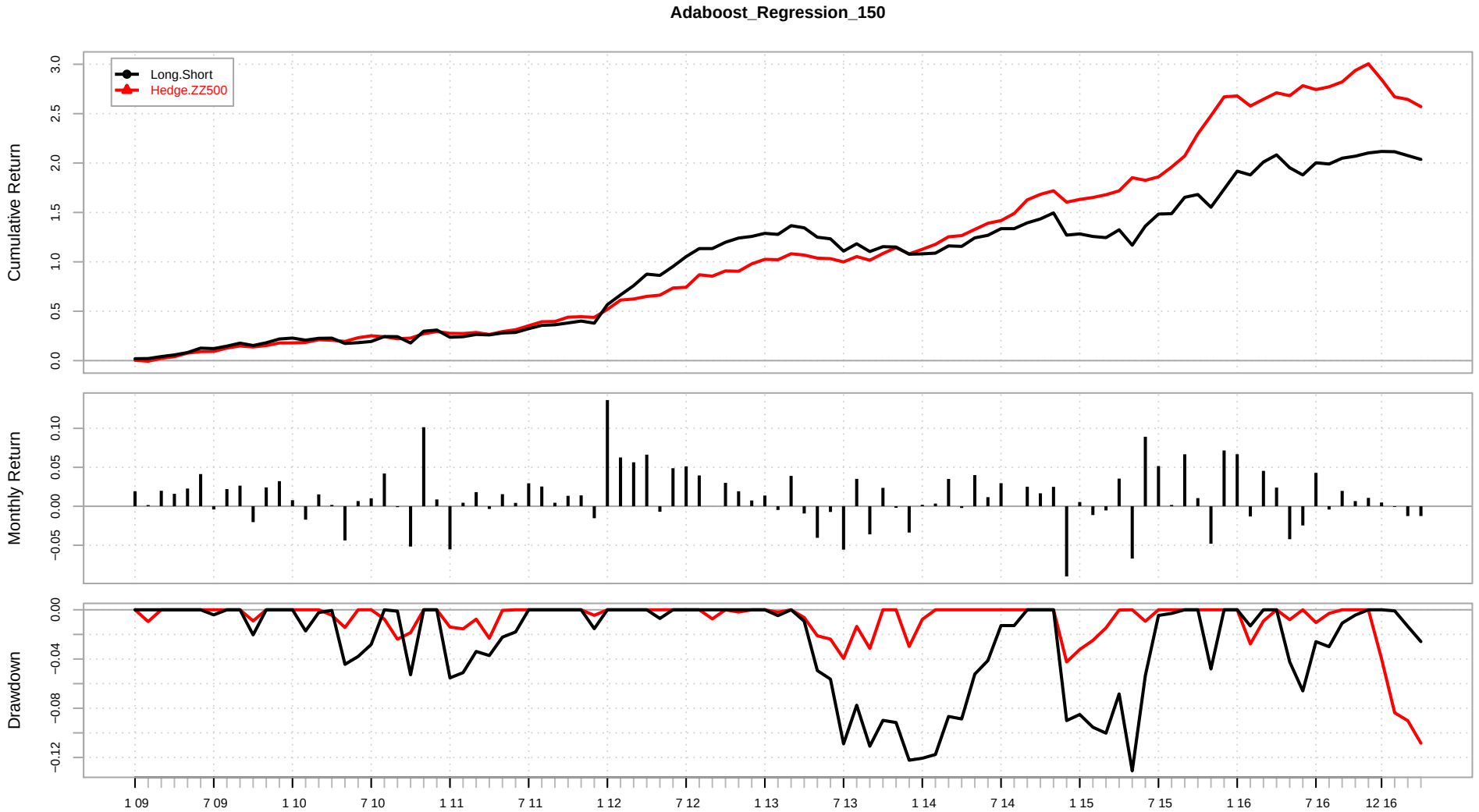
- Ret:16.81%,MaxDrawdown:6.76%,Sharpe ratio:2.274,Score:53.1%.

Adaboost_200_EW



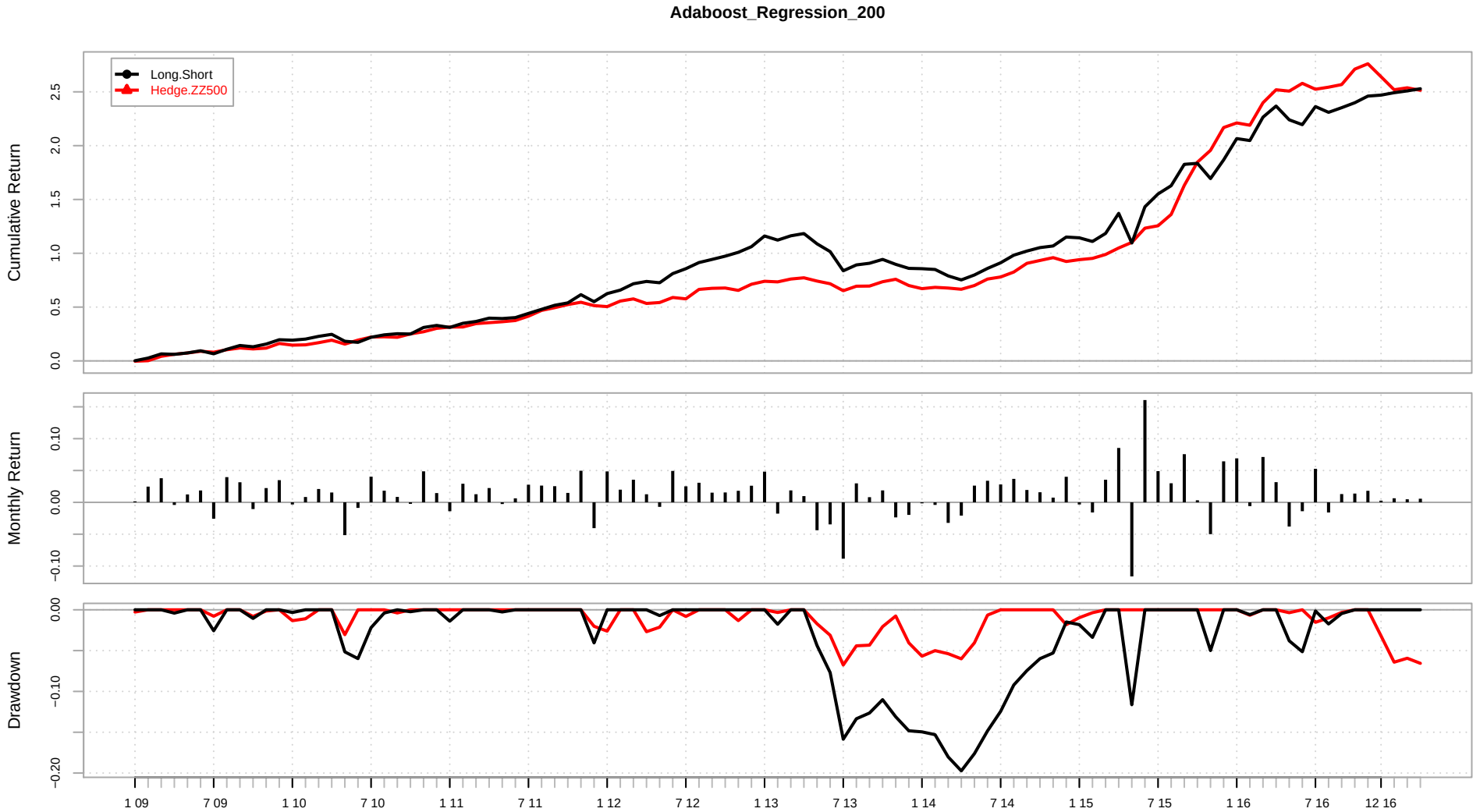
- Ret:16.67%,MaxDrawdown:6.07%,Sharpe ratio:2.268,Score:53.1%.

Adaboost_Regression_150_EW



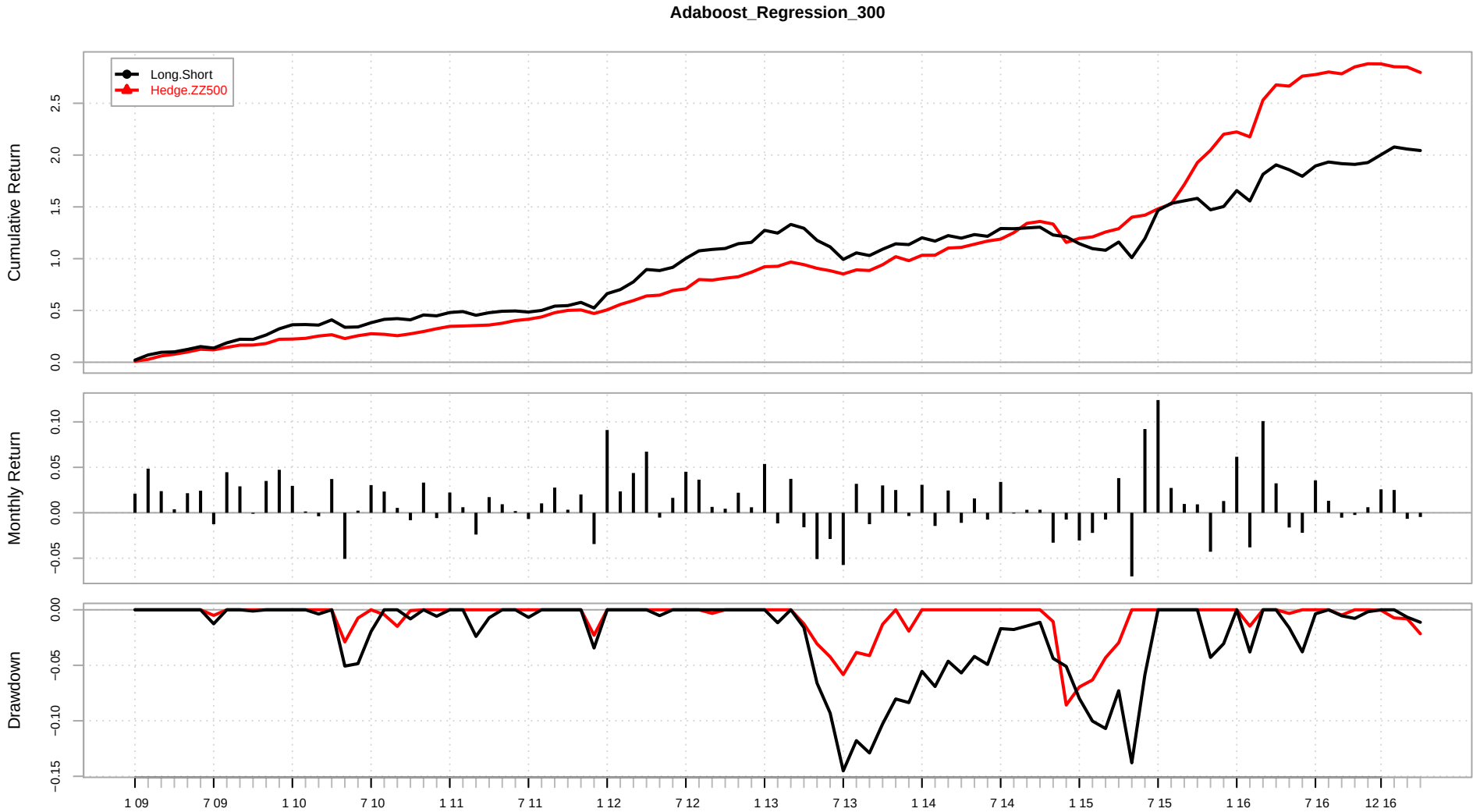
- Ret:16.68%,MaxDrawdown:10.84%,Sharpe ratio:1.953,Score:52.71%.

Adaboost_Regression_200_EW



- Ret:16.45%,MaxDrawdown:6.77%,Sharpe ratio:1.823,Score:52.71%.

Adaboost_Regression_300_EW



- Ret:17.56%,MaxDrawdown:8.58%,Sharpe ratio:2.087,Score:52.71%.