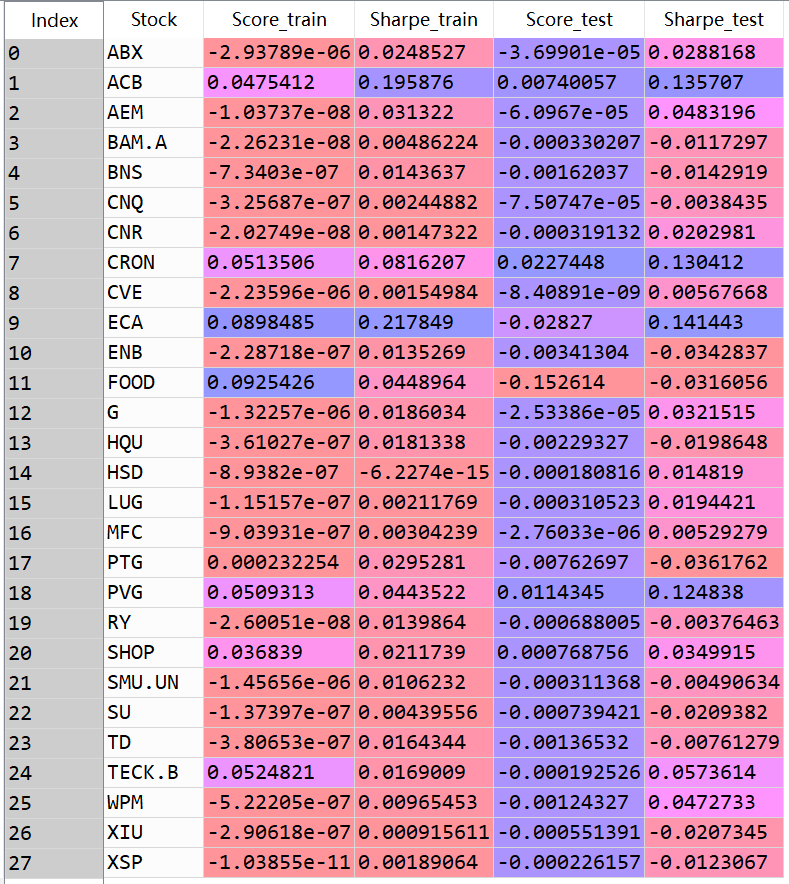
1. RF, n = 100, max\_depth = 2



Select the ones with train score > 0.01:

ACB, CRON, ECA, FOOD, PVG, SHOP, TECK.B are good

1. Look at each’s feature importance, select the ones with score>0.01

ACB, 'mid\_momentum\_20ord', 'transaction\_spread\_5ord', 'volum\_imbalance',

CRON, 'mid\_momentum\_10ord', 'mid\_momentum\_30s', 'mid\_momentum\_50ord',

'smart\_price\_momentum\_50ord', 'smart\_price\_snapshot', 'spread\_10s',

'spread\_30s', 'spread\_5s', 'spread\_diff\_10s', 'spread\_diff\_30s',

'trade\_sign\_10ord', 'trade\_sign\_2s', 'transaction\_spread\_10ord',

'volum\_imbalance'

PVG, 'mid\_momentum\_10ord', 'mid\_momentum\_10s', 'mid\_momentum\_30s',

'mid\_momentum\_50ord', 'mid\_snapshot', 'smart\_price\_momentum\_10ord',

'smart\_price\_momentum\_10s', 'smart\_price\_momentum\_20ord',

'smart\_price\_momentum\_30s', 'smart\_price\_momentum\_50ord',

'smart\_price\_momentum\_5ord', 'smart\_price\_snapshot', 'spread\_10s',

'spread\_30s', 'spread\_50ord', 'transaction\_spread\_10ord',

'volum\_imbalance'

Refit models

* 1. Random Forest

The result seems to be not promising, so we try to tune the parameter.

Unfortunately, as we tune the parameter, we find it extremely hard to improve the performance on test set score. “ACB”, “CRON”, “PVG” always seem to have better performance than the other stocks as the parameter changes.

Final model is:

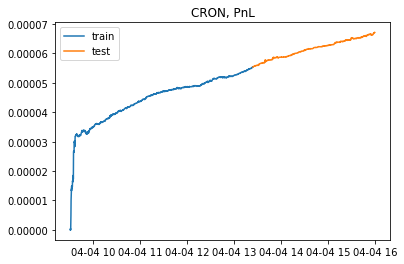
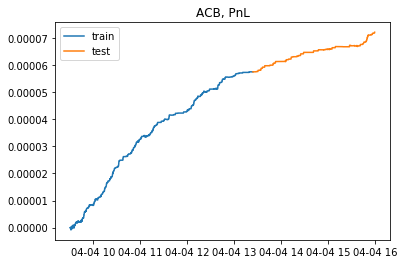
model\_param = {'max\_depth': 2, 'random\_state': 0, 'n\_estimators': 100}

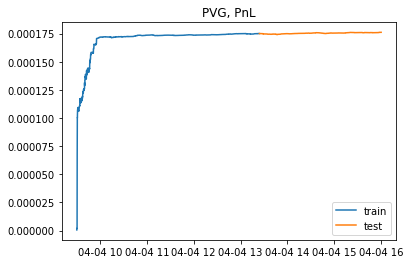
('Train Score: ', 0.04592)

('Train Sharpe: ', 0.19169)

('Test Score: ', 0.00549)

('Test Sharpe: ', 0.1336)





Cron:

('Train Score: ', 0.04081)

('Train Sharpe: ', 0.07381)

('Test Score: ', 0.01828)

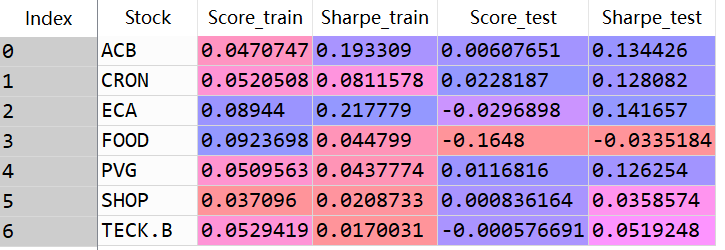
('Test Sharpe: ', 0.14117)

('Train Score: ', 0.04106)

('Train Sharpe: ', 0.04137)

('Test Score: ', -0.00015)

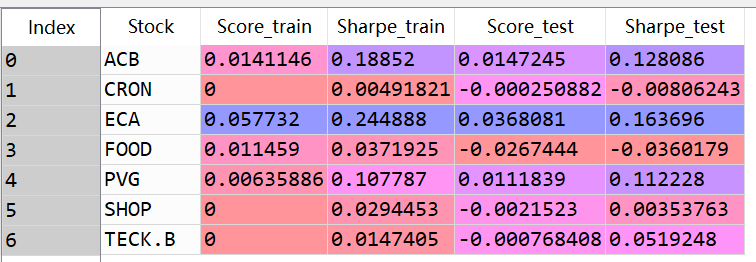
('Test Sharpe: ', 0.02286)



* 1. Elastic Net

At this stage, we can also try other models

We tune the alpha parameter in elastic net and find alpha=1e-6 seems to perform well.



“ACB”, “ECA”, “PVG” are good! Similar but a little different from rf.

* + 1. ACB

Only coefficient for 'volum\_imbalance’ is positive (corresponding to RF)

Select it and use ridge to refit, tune the parameter, choose alpha = 1

ACB:

('Train Score: ', 0.04081)

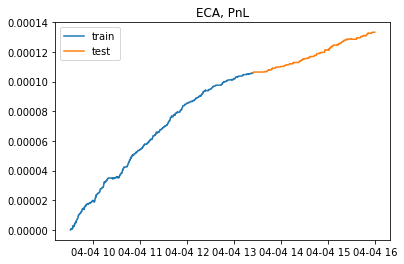
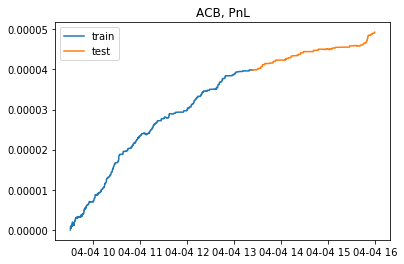
('Train Sharpe: ', 0.19358)

('Test Score: ', 0.03029)

('Test Sharpe: ', 0.13442)

* + 1. ECA

# 'volum\_imbalance' is the only selected factor, again we use ridge to refit it

# and tune the parameter, again alpha = 1 is good

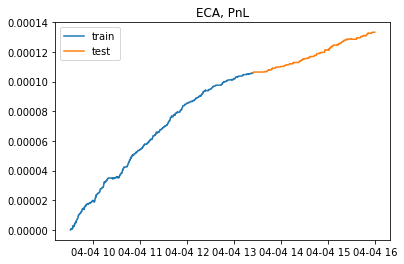
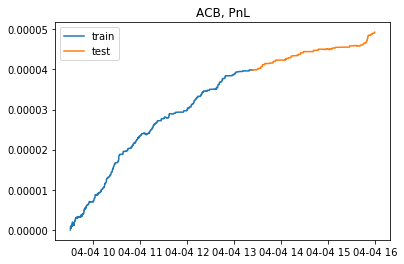
ECA:

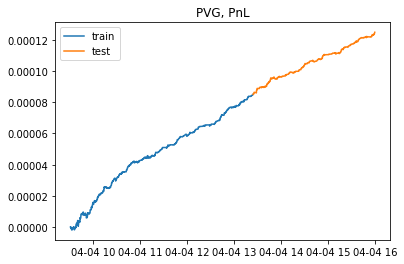
('Train Score: ', 0.05664)

('Train Sharpe: ', 0.24486)

('Test Score: ', 0.03735)

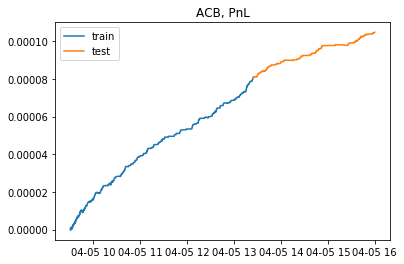
('Test Sharpe: ', 0.16369)





**Test**

ACB, RF



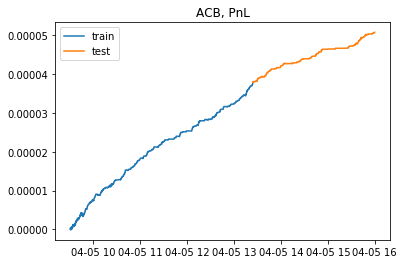
('Train Score: ', 0.04437)

('Train Sharpe: ', 0.1973)

('Test Score: ', -0.00933)

('Test Sharpe: ', 0.16044)

ACB, LR



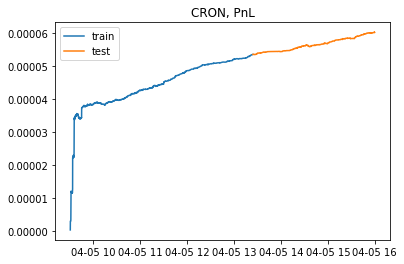
('Train Score: ', 0.02873)

('Train Sharpe: ', 0.19865)

('Test Score: ', 0.03595)

('Test Sharpe: ', 0.17929)

CRON, RF



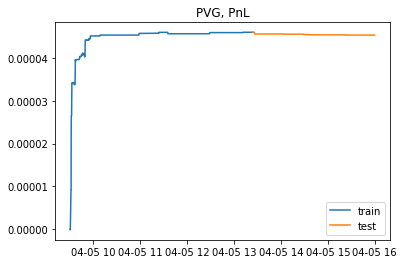
('Train Score: ', 0.04696)

('Train Sharpe: ', 0.05701)

('Test Score: ', 0.00498)

('Test Sharpe: ', 0.13242)

PVG, RF



PVG:

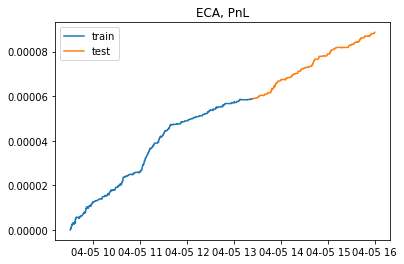
('Train Score: ', 0.0305)

('Train Sharpe: ', 0.05488)

('Test Score: ', -0.00582)

('Test Sharpe: ', -0.03131)

ECA, LR



ECA:

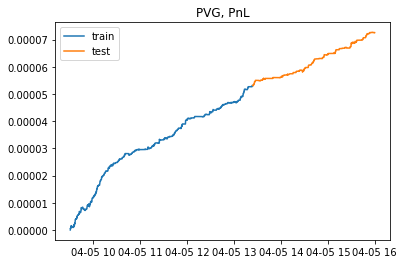
('Train Score: ', 0.04294)

('Train Sharpe: ', 0.19351)

('Test Score: ', 0.05521)

('Test Sharpe: ', 0.18822)

PVG, LR



PVG:

('Train Score: ', 0.03038)

('Train Sharpe: ', 0.17469)

('Test Score: ', 0.00327)

('Test Sharpe: ', 0.13505)