

Back Testing on Investment Strategies

by

Simon Zhang

A report

presented to the Cheriton D. School of Computer Science

and School of Accounting and Finance

in fulfilment of the

course requirement of

CFM 301

Introduction and Motivation

With the help of advanced technologies, investors become more interested on technical analysis on financial investment. There are lots of anomalies becoming more and more interesting to investors; however, some anomalies may not be reasonable indicator for them. According to Harvey, Liu and Zhu (2016), a large number of anomalies that appear in published articles are not performing as expected. In this report, a series of back testing will be applied to some anomalies and true stories of these anomalies will be discovered based on data from past few decades.

Methodology

Data used in this report is retrieved from Wharton Research Data Services. Two data sources are retrieved. First is from Compustat which includes asset, common equity, earning per share, income before extraordinary items and sales; while, second is from CRSP which includes price, shares outstanding, WRDS value-weighted market returns and return on S&P 500 index and return. Besides the common factors listed above, this report also includes two others factors which include share volumes and spread between ask and bid.

Data preparation is important for testing. Since Compustat only provides quarterly data, this report will apply three-month rule. That is, to take into account of reporting time, this report use a three month lag rule; in other words, this report matches price month with quarterly reporting with a three-month apart. For NA values that appear in dataset, a back fill strategy will be applied first for each ticker, and then rows still contain NA values will be removed from this report since there is a probability that the ticker does not have any value during the period.

Third data set used in this report is Fama French which is also retrieved from WRDS. It is used for Fama-MacBeth regressions model.

Factors

Spread

Spread means the difference between ask price and bid price; in other words, it is the gap between short position and long position. Bid Ask Spread is a good indicator for liquidity. A small spread means the gap between short position and long position is not large, and demand and supply are relatively stable. Therefore, the market for the security is liquid. On the other hand, if the gap between ask and bid is high, then the market for the security is not liquid since supply and demand are unbalanced.

According to “Return and Liquidity Relationships on Market and Accounting Levels in Brazil” by Perobelli, Fama and Sacramento (2016), “it was observed that shares that enjoyed greater market risk/expected return traded more and with a higher spread in prices, generating a negative market liquidity premium in the sample and period considered.”

Volume

Volume simply means number of shares traded over a particular period of time, and in this report the period is 1 month. A large trading volume means there are many trades with regard to the security. This can be a good sign or a bad sign for a short period. If there are large number of selling, then the price may decrease for a short period; on the other hand, if there increasing number of investors who want to engage as long position which may lead to

increasing share price for a short period. Therefore, there might be a potential relationship between volume and rate of return.

According to “The Relationship Between Trading Volume and Stock Returns” by Pathirawasam (2011), contemporary trading volume change is positively related with the stock returns.

Market Capitalization

Market capitalization means total market value which equals share price times share outstanding. First, market capitalization tells how much value the public thinks of the company. Second, market capitalization gives an overview of the size of the company. A large market capitalization may show the company is more risk-resistant compared with other companies with low market value.

According to “The Effect of Financial Leverage and Market Size on Stock Returns on the Ghana Stock Exchange: Evidence from Selected Stocks in the Manufacturing Sector” by Acheampong (2014), his study with Ordinary Least Square method demonstrates there is a positive relationship between market size and stock return, although “the size effect within the manufacturing sector was however very limited.”

Price-To-Sales Ratio

The price to share ratio is calculated by taking a company's market value and divide it by the company's total sale or revenue over the past 12 month. In this report, monthly data is used for all companies so there should not be a huge difference. The ratio shows how much the market value every dollar of the company's sale. In this report, the reciprocal of PS ratio will be used as SP ratio because SP ratio means how much sale generated per dollar of market value. If SP ratio is high, then the company has great potential.

According to “Do Sales-Price and Debt-Equity Explain Stock Returns Better than Book-Market and Firm Size?” by Barbee, Mukherji and Raines (1996), there is a strong relationship between sales-to-price ratio.

Earning Per Share

Earning per share (EPS) is pretty similar to sale to price ratio. EPS is defined as a company's profit divided by share outstanding. It demonstrates how much profit for each share, and is a good indicator for corporate value.

According to “The Relationship between Earnings and Stock Returns: Empirical Evidence from the Greek Capital Market” by Dimitropoulos and Asteriou (2009), “the increased ability of the price and return models to explain better the aforementioned relationship by providing highly significant earnings response coefficients.”

Fama-MacBeth regressions

In Fama-MacBeth Regression, cross sectional regressions are applied to each period on factors listed above. The result is shown below.

As can be seen from the figure below (see figure 1), the t-statistics are really low. There are several reasons for this. One of the most important one is data clean. The data clean process in this report may not be good. For example, outlier filter and market capitalization lower bound rules are not applied to the dataset. However, we can still figure out that some factors are more important than other factors. For example, market capitalization has a significant placement compared with other factors involved in this report. Sales-to-Price and earning per share are

relatively more strongly related to rate of return compared with volume. Note that the t-statistic of

```
param_est.mean()
vol      4.357876e-07
spread   -3.310861e-02
sizeln    1.817434e-02
s2p       2.159207e-01
epspxq    1.025397e-02
dtype: float64

param_est.std()
vol      0.000109
spread    1.034140
sizeln    0.148811
s2p       9.605360
epspxq    0.252173
dtype: float64

param_tstat = param_est.mean() / (param_est.std() * np.sqrt(252))

param_tstat
vol      0.000251
spread   -0.002017
sizeln    0.007693
s2p       0.001416
epspxq    0.002561
dtype: float64
```

Figure 1: Fama-MacBeth Regression

spread is negative in this report is negative. This because the spread should be negatively related to the return. For example, a small spread means the market is really liquidity for the security, and the rate of return is reasonably high.

Machine Learning

The report uses three different machine learning models to analyse dataset including simple linear regression model, partial least square regression and elastic net. Code of this part is based on Professor Huang's Machine Learning lecture 1 and lecture 2.

Three regression model results share the same pattern (see figure 2,3,4). However, it is not hard to detect that, for all of the models, regression model results that are derived from the training data set are not fitting the testing data set. This can be supported by R-square value of regression models (see figure 5,6). The R-square value is pretty low in both ElasticNet

Regression and Partial Least Linear Regression Models. The good news is that the mean of the actual data set and the mean of the predicted values are similar and they are gathered at quite similar level which is little bit positive as shown in the graph.

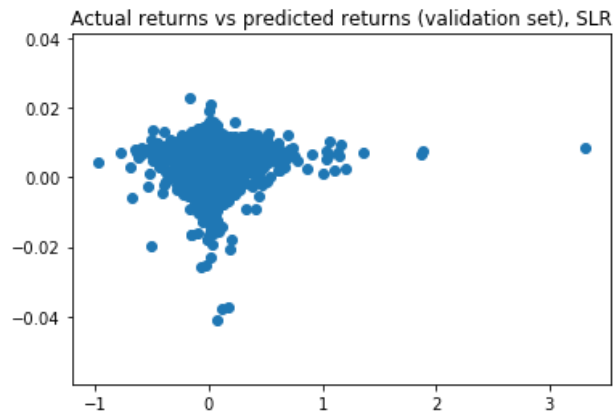


Figure 2: Simple Linear Regression

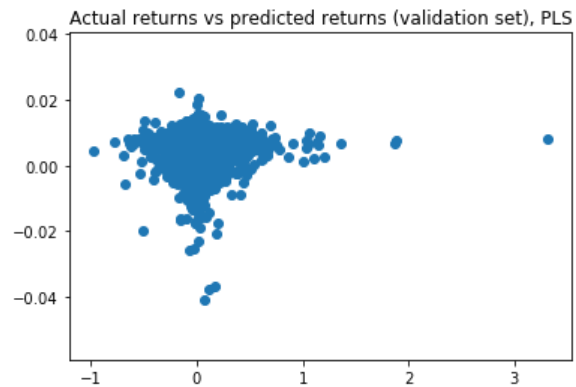


Figure 3: Partial Linear Regression

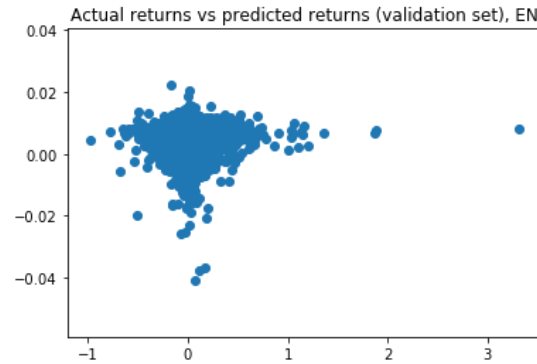


Figure 4: ElasticNet

```
Best Config{'alpha': 0.01, 'l1_ratio': 0.06}
Validation R2: 0.00029800209302122127
```

Figure 5: R^2 of ElasticNet Regression

```
Best config: {'n_components': 1}
Validation R2: 0.0002755757789726365
```

Figure 6: R^2 of Partial Linear Regression

Conclusion:

In this report, five factors are included to back test their relationship with return including spread, volume, market capitalization, price-to-sales ratio and earning per share. These factors are commonly used in different investment strategies, and this report gives an overview of their performance from year of 1980 to 2020.

As can be seen from data, no matter traditional approach like Fama-MacBeth or novel machine learning method, they can give somehow reasonable prediction but not as precise as investors expected. However, it is not hard to admit that the rising of computational finance along with

other assistance will help neutralize future financial market can somehow change investors trading behavior. They are Imperceptibly influencing investors' trading strategies.

Appendix:

Codes included in this report:

Data_pre.ipynb

Data_pre_2.ipynb

In_sample.ipynb

Fm-training.ipynb

ml-datapre.ipynb

ml-training.ipynb

References:

C. Pathirawasam (2011). "The Relationship Between Trading Volume and Stock Returns".

Retrieved from: <https://www.cjournal.cz/files/65.pdf>.

F. Perobelli, R. Famá, L. Sacramento (2016). "Return and Liquidity Relationships on Market and

Accounting Levels in Brazil". Retrieved from:

<https://www.scielo.br/j/rcf/a/hdQcmkFJpnR8ns6Bq4jJ6Zx/?lang=en> .

P. Acheampong (2014). "The Effect of Financial Leverage and Market Size on Stock Returns on the Ghana Stock Exchange: Evidence from Selected Stocks in the Manufacturing Sector".

Retrieved from:

https://www.researchgate.net/publication/272659141_The_Effect_of_Financial_Levera

[ge and Market Size on Stock Returns on the Ghana Stock Exchange Evidence from Selected Stocks in the Manufacturing Sector .](#)

William C. Barbee, Jr., Sandip Mukherji and Gary A. Raines (1996). "Do Sales-Price and Debt-

Equity Explain Stock Returns Better than Book-Market and Firm Size?". Retrieved from:

<https://www-jstor->

[org.proxy.lib.uwaterloo.ca/stable/pdf/4479907.pdf?refreqid=excelsior%3Ad31c4781bf3](https://www-jstor-org.proxy.lib.uwaterloo.ca/stable/pdf/4479907.pdf?refreqid=excelsior%3Ad31c4781bf3)

[22b85748689bb88b319b9&ab_segments=&origin=](https://www-jstor-org.proxy.lib.uwaterloo.ca/stable/pdf/4479907.pdf?refreqid=excelsior%3Ad31c4781bf322b85748689bb88b319b9&ab_segments=&origin=) .

P. Dimitropoulos, D. Asteriou (2009). "The Relationship between Earnings and Stock Returns:

Empirical Evidence from the Greek Capital Market". Retrieved from:

[https://www.researchgate.net/publication/42386600 The Relationship between Earn](https://www.researchgate.net/publication/42386600_The_Relationship_between_Earnings_and_Stock_Returns_Empirical_Evidence_from_the_Greek_Capital_Market)

[ings and Stock Returns Empirical Evidence from the Greek Capital Market .](https://www.researchgate.net/publication/42386600_The_Relationship_between_Earnings_and_Stock_Returns_Empirical_Evidence_from_the_Greek_Capital_Market)