Report v2

Xuewan Zhao

05/12/2019

In this project, we are going to combine fundamental factors and technical factors to construct portfolios. Fundamental factors are used to pick more predictable stocks. Then we would use technical factors to construct forecasting models in trading.

There are two primary methods used to analyze securities and make investment decisions: fundamental analysis and technical analysis. Fundamental analysis involves analyzing a company's financial statements to determine the fair value of the business, while technical analysis assumes that a security's price already reflects all publicly-available information and instead focuses on the statistical analysis of price movements. Technical analysis attempts to understand the market sentiment behind price trends by looking for patterns and trends rather than analyzing a security's fundamental attributes.

Introduction

Fundamental analysis

Fundamental analysis determines the health and performance of an underlying company by looking at key numbers and economic indicators. The purpose is to identify fundamentally strong companies or industries and fundamentally weak companies or industries. Investors go long (purchasing with the expectation that the stock will rise in value) on the companies that are strong, and short (selling shares that you believe will drop in value with the expectation of repurchasing when at a lower price) the companies that are weak. This method of security analysis is considered to be the opposite of technical analysis, which forecasts the direction of prices through the analysis of historical market data, such as price and volume.

For stocks and equity instruments, fundamental analysis uses revenues, earnings, future growth, return on equity, profit margins, and other data to determine a company's underlying value and potential for future growth. In terms of stocks, fundamental analysis focuses on the financial statements of the company being evaluated. One of the most famous and successful fundamental analysts is the so-called "Oracle of Omaha," Warren Buffett, who is well known for successfully employing fundamental analysis to pick securities.

Corporate fundamental data (anything that might be found on a balance sheet), is an incredibly useful source of information. Fundamental data can be used to value companies in pricing models, and one important analysis is how predictive of future returns each fundamental factor is.

In this project, following fundamental factors are considered. The factors in this section are discussed in Chapter 5 of Quantitative Equity Portfolio Management by Qian, Hua and Sorensen, Chapman and Hall, 2007.

- Market Value
- Enterprise Value
- Cash Dividend
- Earnings before Interest, Taxes, Depreciation and Amortization to Enterprise Value
- Trailing 12-month earnings to market capitalization
- Book to market capitalization
- Sales to Enterprise Value

Technical analysis

Technical analysis is a trading discipline employed to evaluate investments and identify trading opportunities by analyzing statistical trends gathered from trading activity, such as price movement and volume. Unlike fundamental analysts, who attempt to evaluate a security's intrinsic value, technical analysts focus on patterns of price movements, trading signals and various other analytical charting tools to evaluate a security's strength or weakness. Technical analysis can be used on any security with historical trading data.

Across the industry there are hundreds of patterns and signals that have been developed by researchers to support technical analysis trading. Numerous types of trading systems are also developed to help them forecast and trade on price movements. Some indicators are focused primarily on identifying the current market trend, including support and resistance areas, while others are focused on determining the strength of a trend and the likelihood of its continuation. Commonly used technical indicators and charting patterns include trendlines, channels, moving averages and momentum indicators.

In general, technical analysts look at the following broad types of indicators:

- price trends
- moving averages
- volume and momentum indicators
- oscillators
- support and resistance levels

Strength and weakness

There are many studies focus only on one of fundamental analysis and technical analysis. Fundamental analysis serves for long-term value investment while technical analysis serves for short-term trading.

However, as fundamental analysis focus on long-term performance of the company, it can't reflect whether we should entry this company at current price. While technical analysis focus on short-term trading, if we could choose a fundamental sound instrument, the probability of success could be improved.

By using both fundamental and technical analysis, the probability of success could be improved. The idea is to use fundamental analysis to select sound candidates and use technical analysis to determine the ideal entry/exit points.

Data

Data source

Data used in this project contains historical S&P 500 stocks from 1998/01/01 to 2018/12/31. This time period covers Dot-com bubble, 911 attacks, financial crisis, etc. Thus, our research would be more reliable and makes it easy to see its performance in extreme situations.

Data source for this project is CRSP/Compustat Merged Database available from Wharton Research Data Service (WRDS).

Fundamental Factors

Fundamental factors discussed in this project are based on *Value Factors Do Not Forecast Returns for S&P 500 Stocks*.

Fundamental factors are summarized in Table 1.

Table 1: Value Factor Description

Value factor	Description
MV	Market Value
EV	Enterprise Value
CD	Cash Dividend
EBITDA2EV	Earnings before Interest, Taxes, Depreciation and Amortization to Enterprise Value
E2PFY0	Trailing 12-month earning to market capitalization
B2P	Book to market capitalization
S2EV	Sales to Enterprise Value

The correlations between the value factors are examined for all stocks in the S&P 500 universe throughout the back-test period. In linear models, a factor may be omitted from the linear regression if it is highly correlated with another factor. Avoiding highly correlated factors could also help avoid multi-collinearity, which results in inaccurate ordinary least squares regression results.

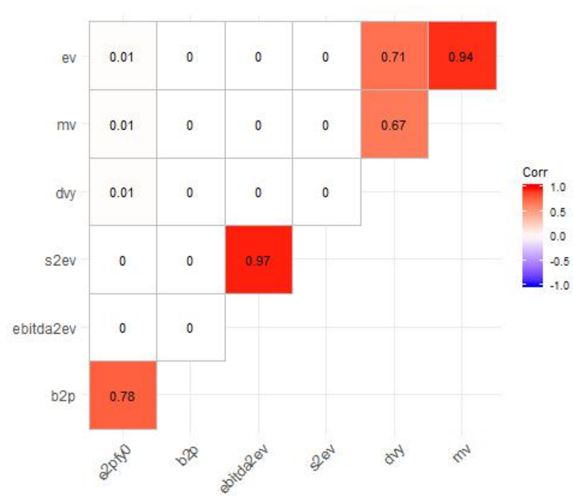

Figure 1: Factor Correlations

As we can see from the heat map, two pairs (EV and MV, S2E and EBITDA2EV) are highly correlated, which makes sense in the stock market.

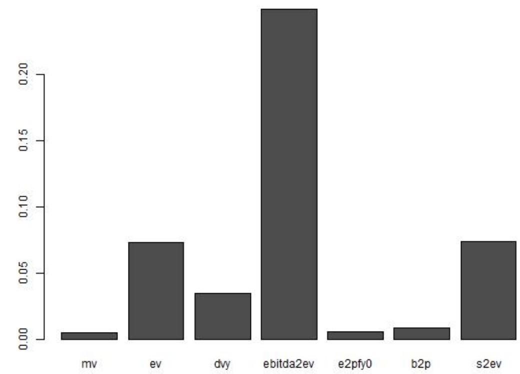

Figure 2: Factors' NA Percentages

As a result, we would omit EV in the following analysis in favor of the MV factor. EBITDA2EV would also be omitted in favor of S2EV.

S&P 500 constitution

The constitution of S&P 500 stocks keeps changing, as companies would enter and exit the stock market. In our model, we are going to use historical S&P 500 constitutions, which could help avoid survivor bias. Some stocks do not exists anymore and the data is missing in CRSP/Compustat database. However, as we can see below, most stocks are still recorded and available.

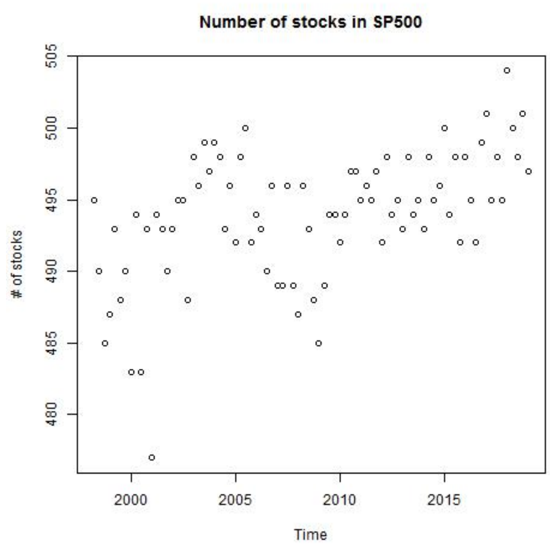

Figure 3: Number of S&P 500 stocks in each quarter

As we can see, as time goes, the number of available stocks increases. However, from 2006 to 2009, due to financial crisis, many companies defaults, which made the number of available stocks during this period decreases.

Model

Fundamental factor model

We are going to use fundamental factors to select stocks. There are two straight ways to select stocks: 1. Pick stocks by ranking the stocks on the basis of the fundamental factors (one at a time). 2. Pick stocks by ranking the stocks on the return predicted by a linear model constructed from the value factors.

After selecting the stocks, there are also two straight ways to trade in the market: 1. Long the top stocks only. 2. Long the top stocks, short the bottom stocks which makes us market(dollar) neutral.

Ian L. Kaplan has tested these two ways in *Value Factors Do Not Forecast Returns for S&P 500 Stocks*, there are some interesting conclusions could be referred to here. 1. Multi-factor ranking model performs better than single factor model. 2. Long/Short Portfolio performs better than long only portfolio.

Thus, in this project, we are going to use linear multi-factor models to predict quarterly stock return. Then we would pick top and bottom 20 percent stocks to construct the portfolio. We are going to long the stocks which we expect would have high return in next quarter and short the stocks expected to have low returns. Meanwhile, the benchmark would be the portfolio which invests in all available stocks with equal weights.

An important thing needs to be noticed is the releasement date of quarterly fundamental data would be two quarters' later than the quarter fundamental data belongs to. (i.e. the fundamental data for March 1998 would not be released before September 1998.)

Assume current time spot is *t*, given previous quarterly fundamental factors in *t*, we are going to predict the quarterly return at time *t+1*. However, if we take the lag of information into consideration, we are actually using data in *t-3* to predict the quarterly return of *t+1*.

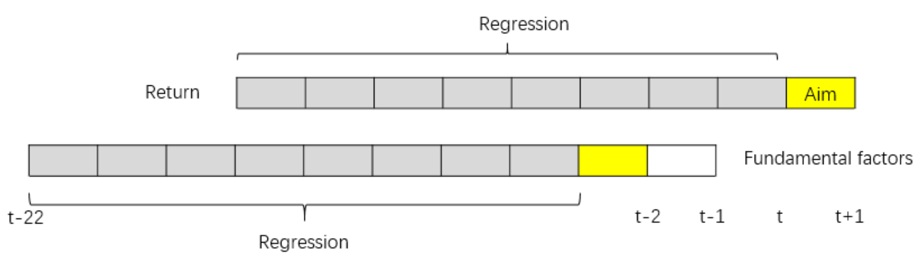

Figure 4: Regression model illustration

As we now have 5 factors, math illustration of regression would be:

$$r_{t+1} = \beta_0 + \beta_1 f_{t,1} + \beta_2 f_{t,2} + \beta_3 f_{t,3} + \beta_4 f_{t,4} + \beta_5 f_{t,5} + \epsilon_t$$

To make our model estimations updated to most recent situation for each time, we are going to use multiple linear regression with rolling window to determine the estimations at time t.

$$\begin{pmatrix} r_t \\ r_{t-1} \\ \vdots \\ r_{t_1 9} \end{pmatrix} = \begin{pmatrix} 1 & f_{t-1,1} & \cdots & f_{t-1,5} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & f_{t-20,1} & \cdots & f_{t-20,5} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_5 \end{pmatrix} + \begin{pmatrix} \epsilon_{t-1} \\ \epsilon_{t-2} \\ \vdots \\ \epsilon_{t-20} \end{pmatrix}$$

After determine the estimations at time t using last 20 quarters' (5 year) data, we are going to forecast the return during time period *t* to *t+1*.

Here could be a topci further study about regression method optimization. For example, we can test to use Robust OLS, WLS instead of OLS, and compare the results. In weighted least squared model, return can be used as the weight. Stocks predicted to have higher return would be given higher weight.

Technical factor model

In this factor model, only technical factors would be used to predict the daily log return of stocks.

Table 2: Technical Factors

Symbol	Variable	Function
SMA20	Simple moving average in 20 days	$\frac{C_t + \cdots + C_{t-n+1}}{n}$
EMA20	Exponentially weighted moving average in 20 days	$kC_t + (1-k)EMA_{t-1}$
Vol	Volume	V_t
MMT	Momentum	$C_t - C_{t-n}$
SKP	Stochastic K%	$100\frac{C_t - LL_{t,t-n}}{HH_{t,t-n} - LL_{t,t-n}}$
SDP	Stochastic D%	$\frac{\sum_{i=0}^{n-1} SKP_{t-i}}{n}$
RSI	Relative Strength Index	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} Up_{t-i})/(\sum_{i=0}^{n-1} Dw_{t-i})}$
MACD	Moving Average Convergence Divergence	$MA(m_1) - MA(m_2)$
LWR	Larry Williamś R%	$\frac{H_{t-n} - C_{t-n}}{H_{t-n} - L_{t-n}}$
ADI	Accumulation Distribution Oscillator	$ADI_{t-1} + \frac{2C_t - H_t - L_t}{H_t - L_t} V_t$
CCI	Commodity Channel Index	$\frac{(H_t + L_t + C_t) - 3SMA_t}{0.045AD_t}$

C: Close price; H: High price; L: Low price; n: Look-back period; HH: Highest high price; LL: Lowest low price; AD: Average deviation.

Here could be a topic further study about other machine learning methods which could be used to construct model with these features. For example, artificial neural network, support vector machines with polynomial and radial basis function kernels.

Stock screening

In this part, we are going to select 40 stocks which have best/worst performance according to our fundamental factor model and examine the performance of our model.

Survivorship bias is the tendency to view the performance of existing investments in the market as a representative comprehensive sample. Survivorship bias can result in the overestimation of historical performance and general attributes of an investment. Survivorship bias may also be known as “survivor bias.”

In this paper, as the constitution of S&P 500 is changing (Some companies would exit the market), there would be survivor bias if we use all stocks in the constitution as our stock pool. To avoid survivor bias, we are going to randomly select 2/3 available stocks as our base stock pool in each quarter.

In this part, we'll use 1998-2008 data as in-sample period and 2009-2018 data as out-of-sample period. As financial crisis happened during 2007-2008, we are not going to separate these two years.

Survivorship bias is the tendency to view the performance of existing investments in the market as a representative comprehensive sample. Survivorship bias can result in the overestimation of historical performance and general attributes of an investment. Survivorship bias may also be known as "survivor bias."

In sample test:

Due to the property of our model, we need at least 5 years' data lookback period. Thus, the earliest quarter could be predicted is first quarter of 2003.

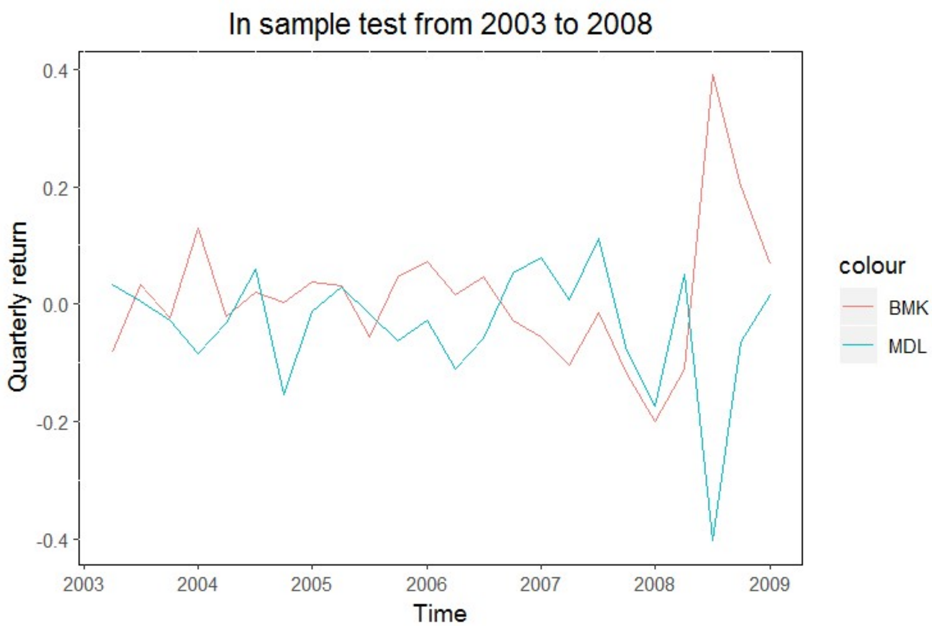

Figure 5: In sample quarterly return

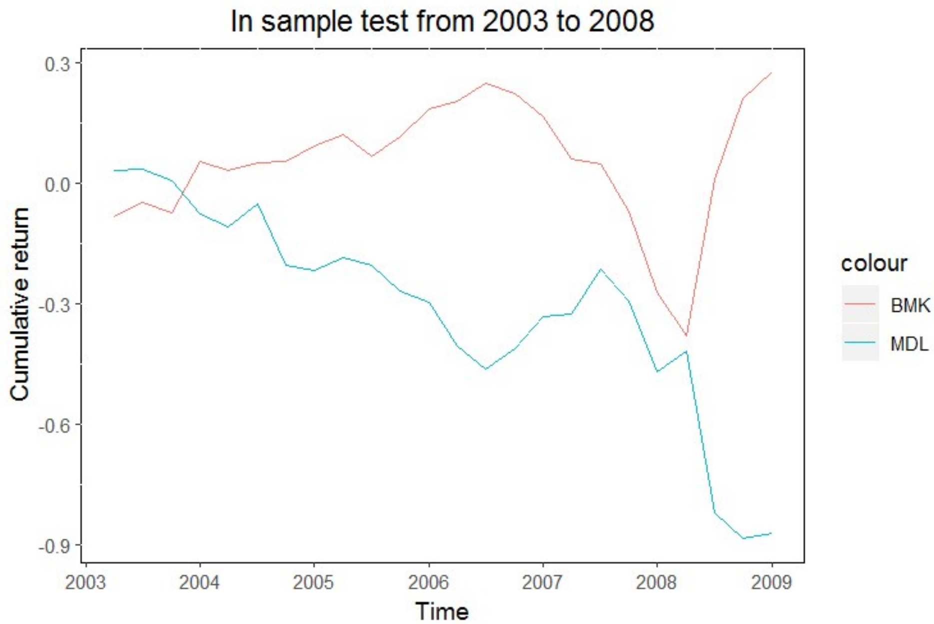

Figure 6: In sample cumulative return

It's interesting to see that the performance of our model always goes opposite with the benchmark. This observation is also supported by R-squared:

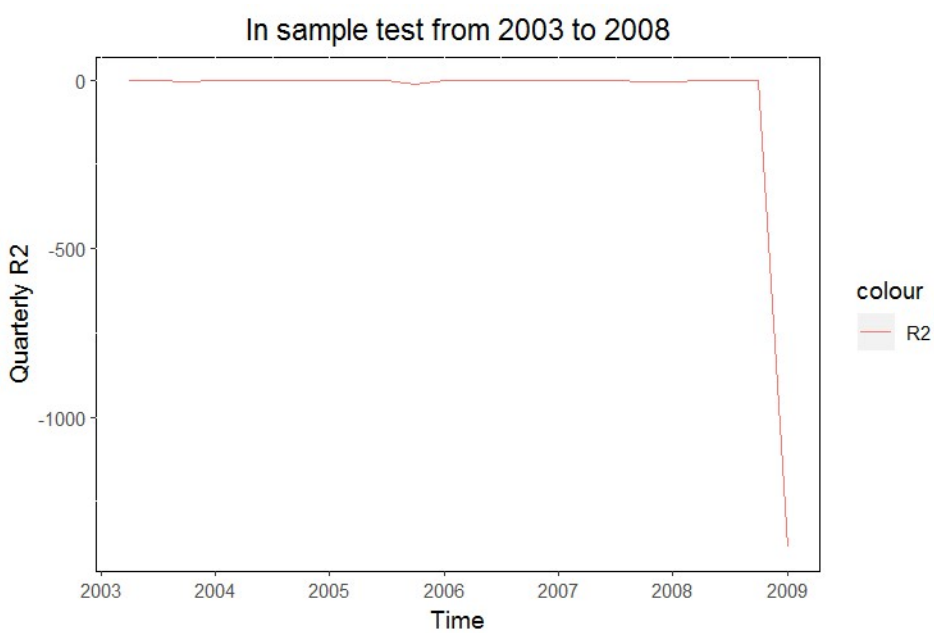

Figure 7: In sample R-squared

We can see, the r-square for each quarter is negative. Even that's the case, we can take advantage of our in-sample test. In out-of-sample test, we are going to assume the stocks would tend to perform in opposite direction with our prediction. Then we would do the opposite trading: long stocks with low expected returns and short stocks with high expected returns.

Another thing needs to be mentioned is: as we can see from results, the market volatiles during and after financial crisis period.

There are some possible reasons why our prediction goes in the opposite direction against real situation:

1. Information already been realized in market. As we are trying to predict the return two quarters' later, although the official reports are just released, the information in official reports may have already been acknowledged by the market through observations or other methods. Thus, the price movements indicated by previous quarterly reports are already realized.
2. The price movements are mean reverting. This is consistent with the speculation above. This would explain why our r-square is always negative.

Verification: We can verify our speculations by: 1. Use the most recent fundamental factors to predict the return in next quarter, i.e., ignoring the 'two quarter's lateness' in real world.

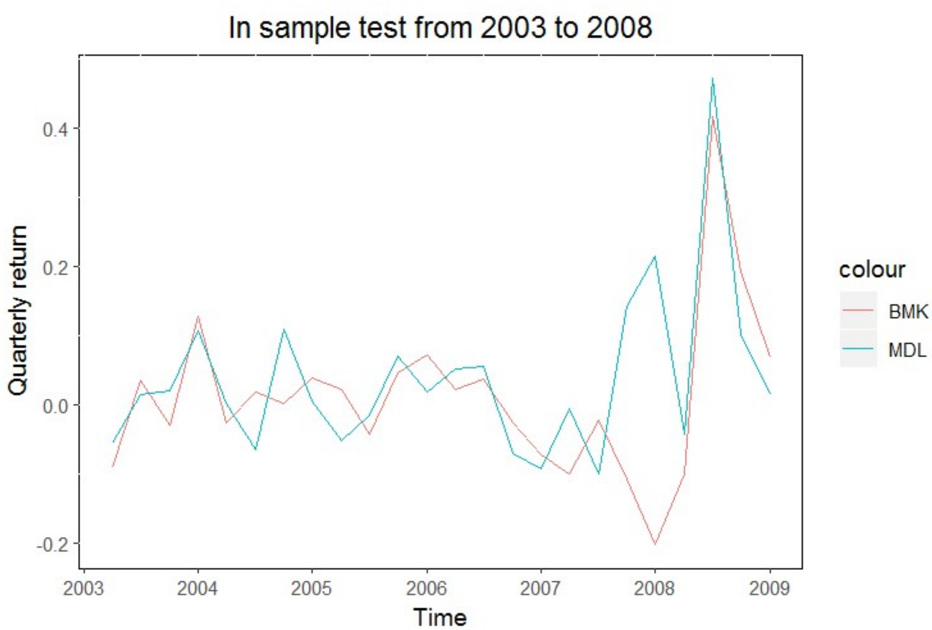

Figure 8: In sample cumulative return v1

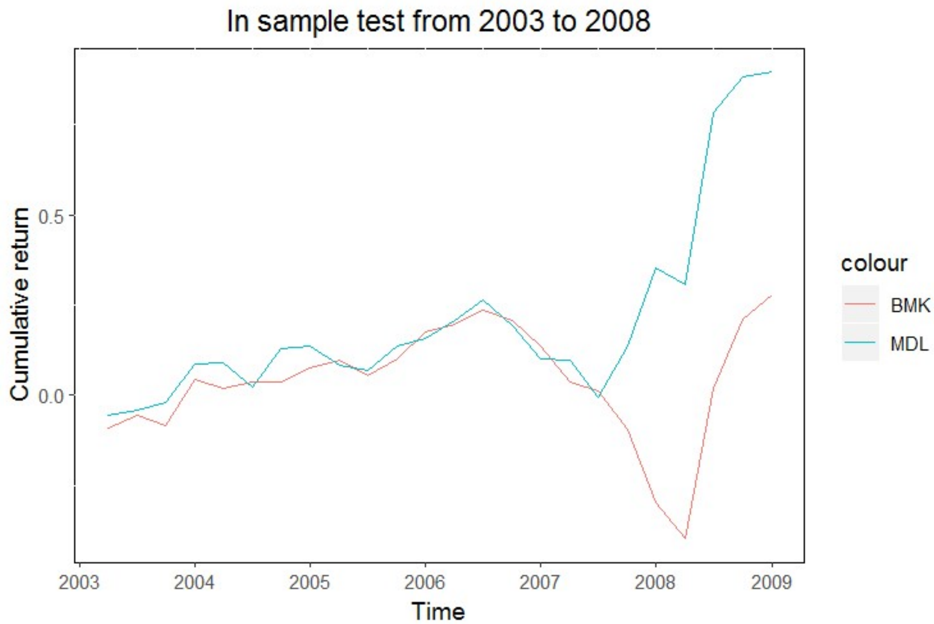

Figure 9: In sample cumulative return v1

Once we changed our trading method and re-test the in-sample period, we'll see this model performs pretty well.

2. Test the mean reverting property of stock returns.

We do the augmented Dickey–Fuller test (ADF) to test the mean reversion property of stock returns. In this part, we collect the p-value of ADF statistics. Once p-value is smaller than 0.05, we recognize this stock as mean reverting. As there are different lags, we would present the percentage of mean reverting stocks with different each lag terms.

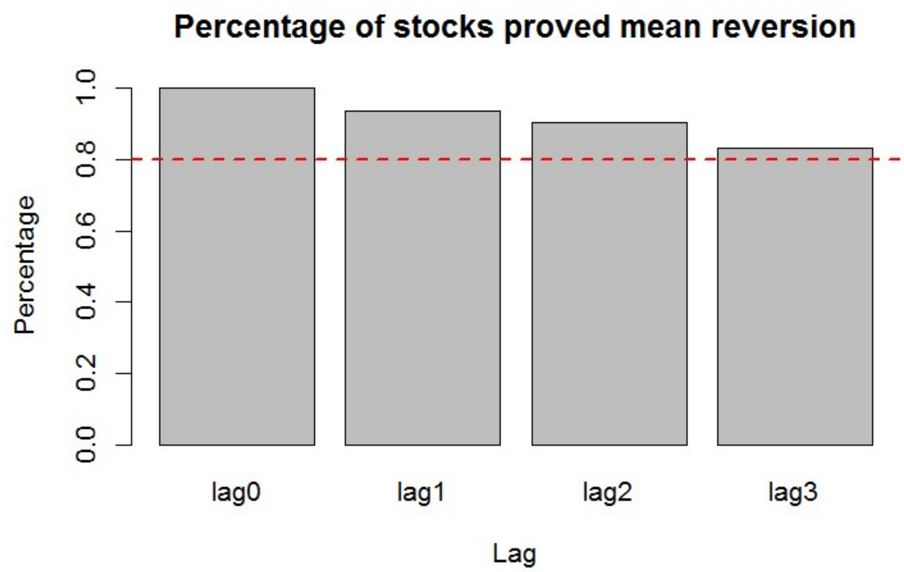

Figure 10: Percentage of mean reverting stocks

Out of sample test:

Now we apply our model to out-of-sample period and check its performance.

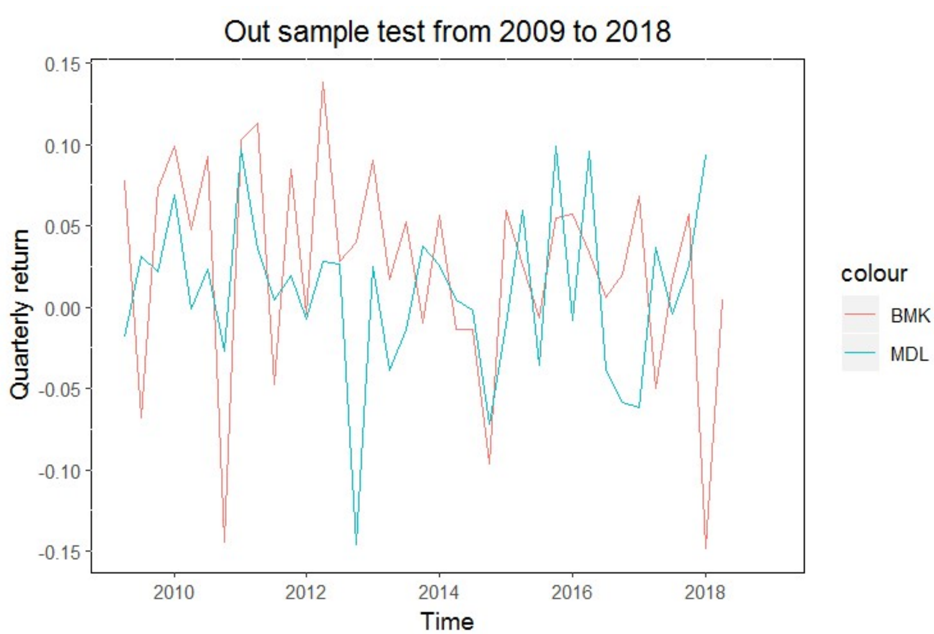

Figure 11: In sample quarterly return

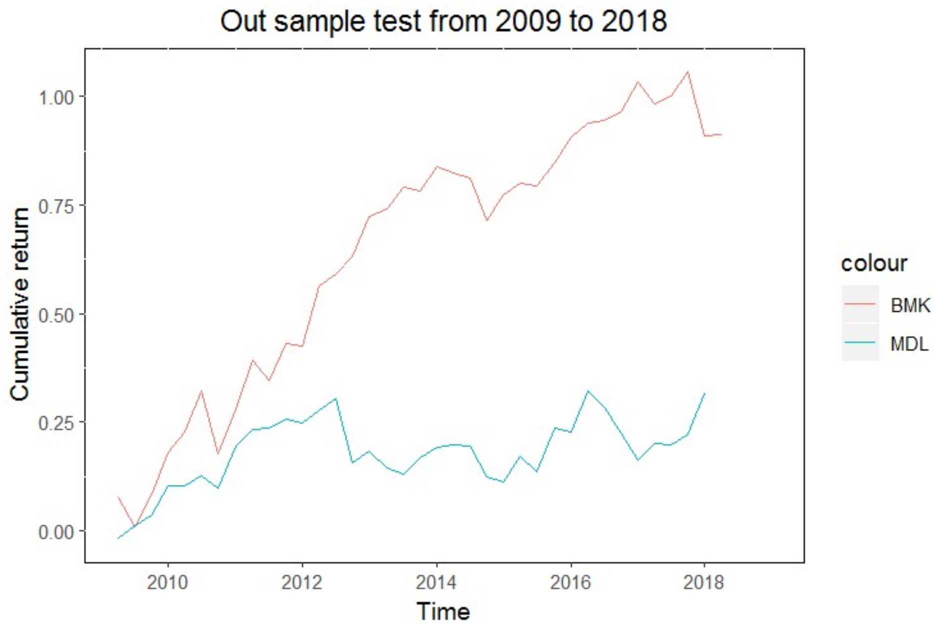

Figure 12: In sample cumulative return

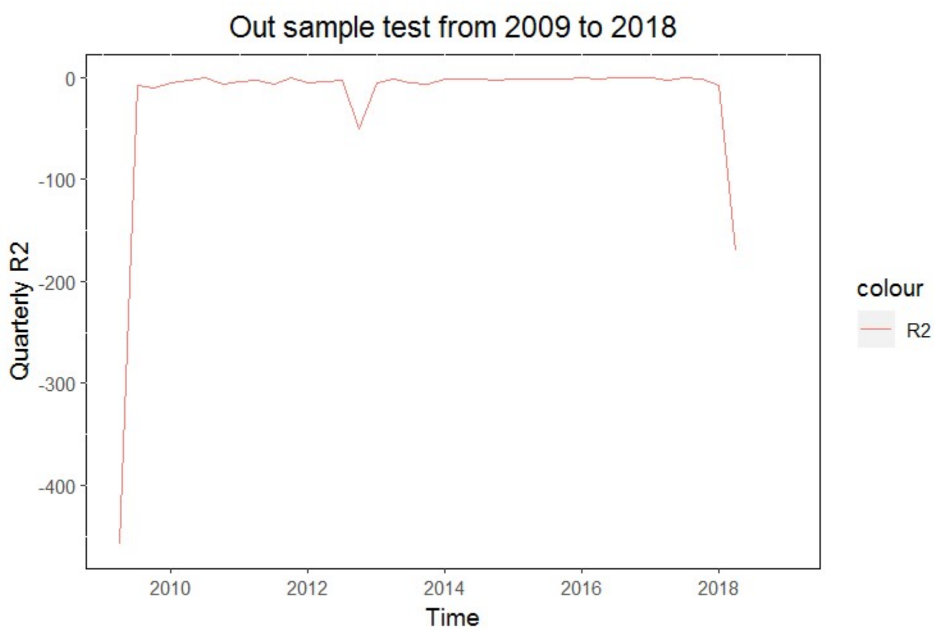

Figure 13: In sample R-squared

We can see that our portfolio generates less return, while its performance is less volatile. We can check the information ratio of our model and the benchmark portfolio.

Information Ratio	Benchmark	Model
In sample (2003-2008)	0.09236724	0.3858518
Out sample (2009-2018)	0.3303293	0.2384175

Table 3: Comparasion of information ratio

If we refer to the 10-year performance of portfolios in the market, we can see our model beats 75% portfolios in the market. However, it can't beat benchmark portfolio during out-of-sample period. There may be some possible reasons:

1. In bull market, most stocks are increasing, only pick top/bottom stocks are not suitable. For example, once we predict all stocks' price would increase, we would not short any stocks.
2. This strategy suits in bear market, i.e., it would be with low risk meanwhile with low return. This speculation is consistent with its low volatility property, at the same time its information ratio is still good.

Verification:

We can verify our speculations by: 1. Verify the performance of stocks during 2009-2018.

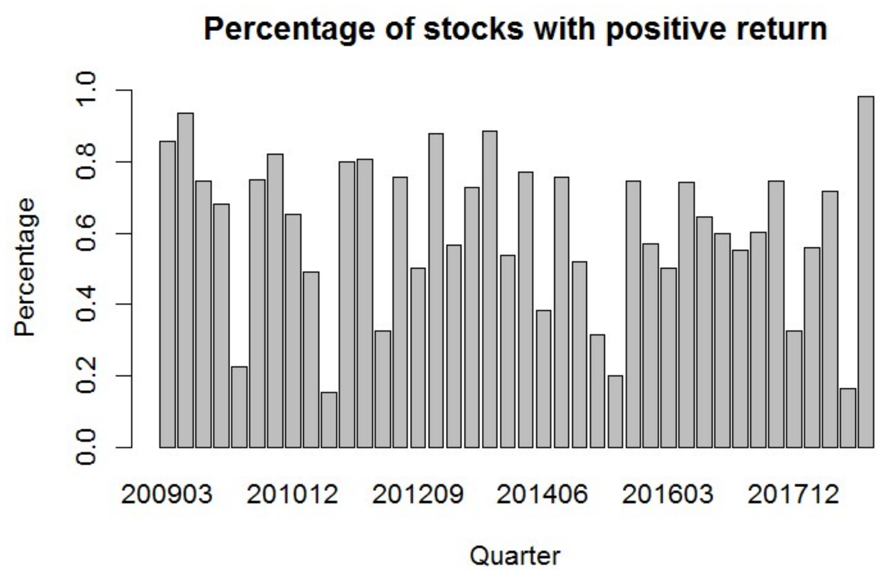

Figure 14: In sample R-squared

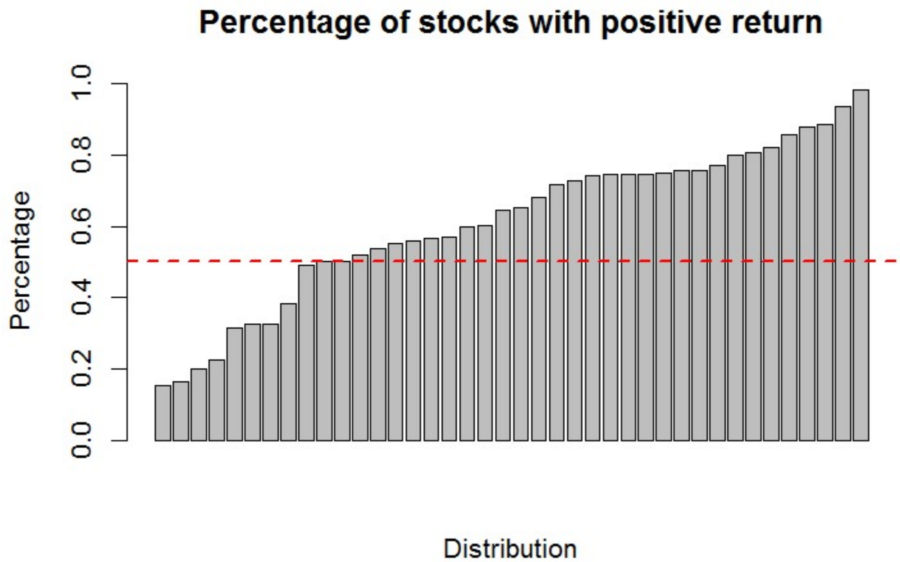

Figure 15: In sample R-squared

We can see clearly from the distribution of the percentage of stocks which have positive return in each quarter, from 2009 to 2018, the whole market status should belong to bull market, which consists with our speculation.

Portfolio trading

As we have tested the performance of the strategy based on fundamental factors, we are going to test the performance of strategies based on technical factors. In this part, we will base on the results from above parts, i.e., we will use stocks selected by fundamental analysis.

We are going to use technical factors to predict daily return of stocks. Once we predict certain stock would have positive return, we will long this stock, vice versa.

$r_{t+1} = \beta_t + \beta_o open_t + \beta_c close_t + \beta_l low_t + \beta_h high_t + \beta_v vol_t + \beta_s sma_t + \beta_e ema_t + \beta_m mmt_t + \beta_{fk} fastK_t + \beta_{fd} fastD_t + \beta_{sd} slowD_t + \beta_r rsi_t + \beta_{macd} macd_t + \beta_{lwr} lwr_t + \beta_a adi_t + \beta_c cci_t$

To analyze the performance of technical factor model itself, we will compare the portfolio return with another benchmark portfolio. Benchmark portfolio would only long or short stocks in each quarter, which means there won't be any daily adjustment on stock holding during a certain quarter period.

We test this model on the time period 2009 to 2018.

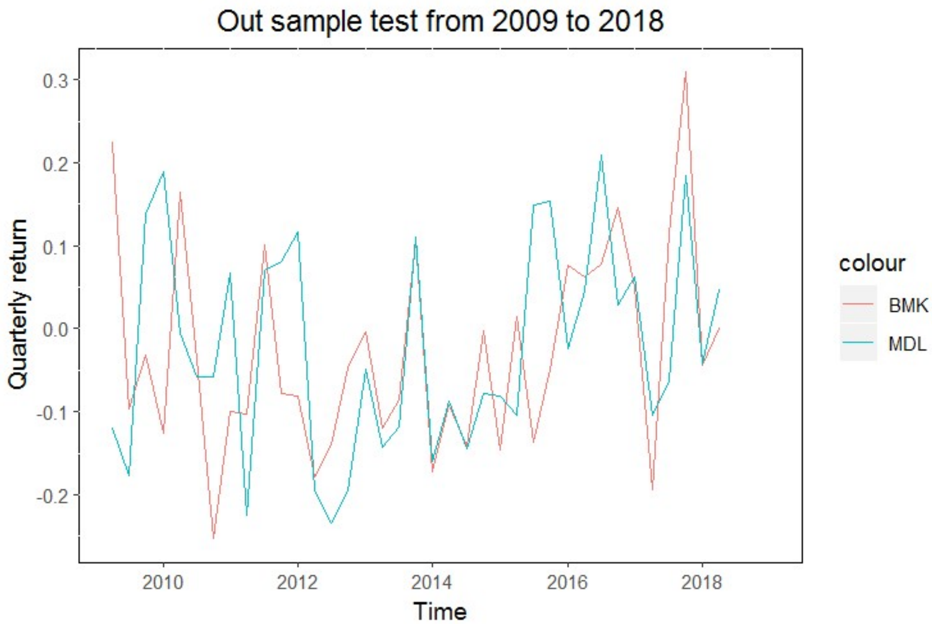

Figure 16: Out of sample quarterly return of technical factor model

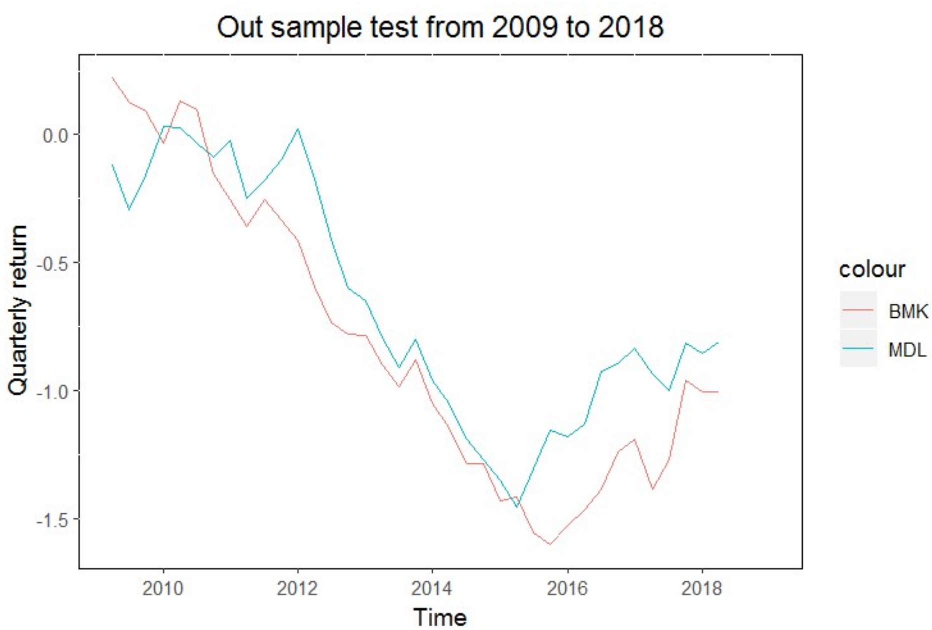

Figure 17: Out of sample cumulative return of technical factor model

As we can see, though we are getting negative return, technical factor model itself improves the performance of the portfolio selected by fundamental factor model.

Conclusion

In this project, we get the following conclusions:

1. Fundamental factor model could generate good information ratio.
2. Fundamental factor model suits in bear market, i.e., it would be with low risk meanwhile with low return.
3. Technical factor model does improve the performance of portfolio constructed based on fundamental factor model.

Another thing needs to be mentioned is, though offical fundamental data would only be released two quarters later than referring quarter, the performance of companies in referring quarter could be analyzed based on other information sources in market. Thus, there would be late efficiency effect if we keep using two quarters' earlier fundamental factors to predict current quarter's performance.

References

1. Kaplan, I., 2014. Value Factors Do Not Forecast Returns for S&P 500 Stocks. Available at SSRN 2407303.
2. Bohl, L., 2017. Stock Forecasting Fundamental Technical Factors.
3. Zephyr. Information Ratio Retrieved from http://www.styleadvisor.com/resources/statfacts/information-ratio