

## Capstone project

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.

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

Table 1 Value Factor Description

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.

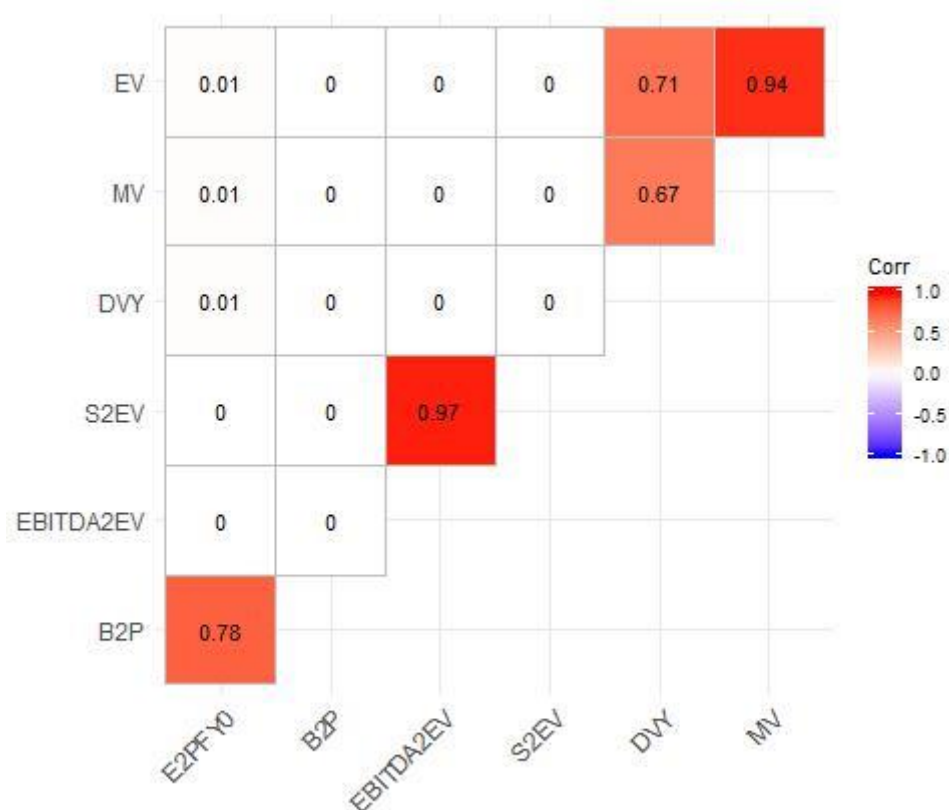


Figure 1 Factor Correlations

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.

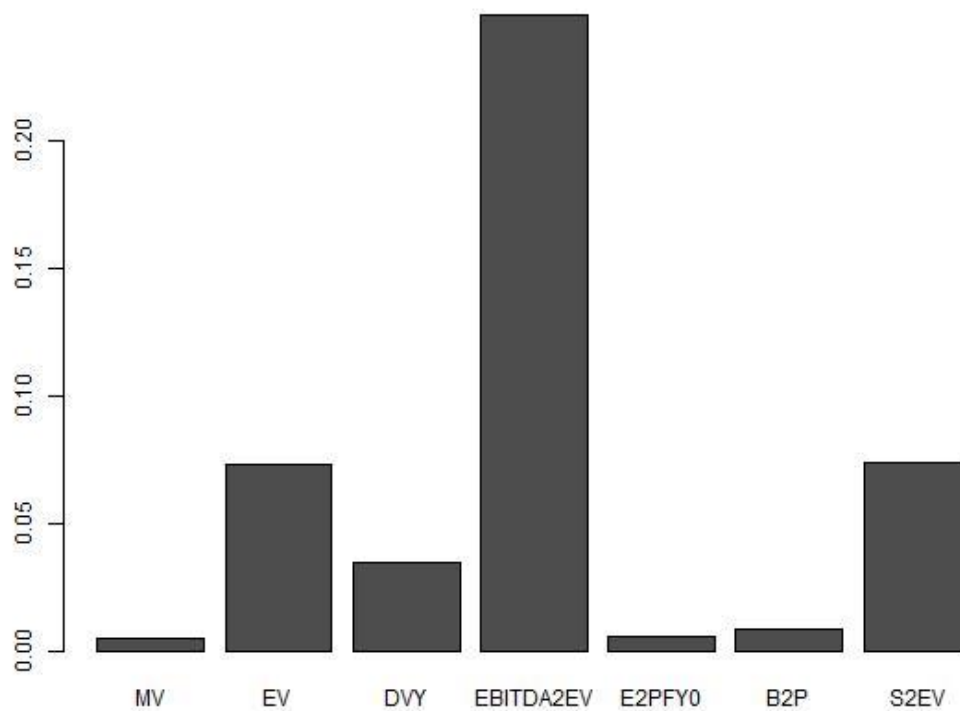


Figure 2 Factors' NA Percentages

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.

## 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. Single factor ranking model performs worse than multi-factor model.
2. Long only portfolio performs than Long/Short 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 10 percent stocks to construct the portfolio.

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  $t+1$ .

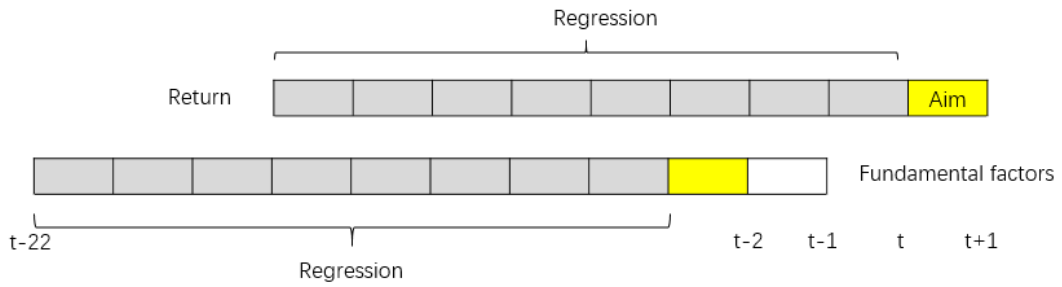


Figure 3 Regression model illustration

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

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

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  $t$ .

$$\begin{pmatrix} r_t \\ r_{t-1} \\ \vdots \\ r_{t-19} \end{pmatrix} = \begin{pmatrix} 1 & f_{t-1,1} & \cdots & f_{t-1,5} \\ \vdots & \vdots & & \vdots \\ 1 & f_{t-20,1} & \cdots & f_{t-20,5} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_5 \end{pmatrix} + \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \\ \varepsilon_{t-20} \end{pmatrix}$$

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  $t+1$ .

To be continued: OLS vs Robust OLS vs WLS

## Technical factor model

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

Symbol	Variable	Function
SMA20	Simple moving average in 20 days	$\frac{C_t + \dots + C_{t-n+1}}{n}, n=20$
EMA20	Exponentially weighted moving average in 20 days	$C_t * k + EMA_{t-1} * (1 - k)$
Vol	Volume	$V_t$
MMT	Momentum	$C_t - C_{t-n}$
SKP	Stochastic K%	$\frac{C_t - LL_{t,t-n}}{HH_{t,t-n} - LL_{t,t-n}} * 100$
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	$MA(m1) - MA(m2)$
LWR	Larry William's R%	$\frac{H_{t-n} - C_{t-n}}{H_{t-n} - L_{t-n}}$
ADI	Accumulation Distribution Oscillator	$ADI_{t-1} + \frac{2 * C_t - H_t - L_t}{H_t - L_t} * V_t$
CCI	Commodity Channel Index	$\frac{(H_t + L_t + C_t) / 3 - SMA_t}{0.015 AD_t}$

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

Table 2 Technical Factors

To be continued: Try to use return as weight? Or equal weighted?

artificial neural network, support vector machines with polynomial and radial basis function kernels.

### **Stock screening**

### **Portfolio trading**

### **References**

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5685607/#pone.0188107.ref002>  
<https://arxiv.org/ftp/arxiv/papers/1801/1801.01777.pdf>