Academic Research Review

Mean Reversion - 4-Factor Model and Overnight Returns

ABSTRACT

This document reports a mean reversion strategy using the 4-factor model on the overnight returns in a universe of correlated instruments, with the example used here being the NYSE. This report is meant to be readable for the average trader and investor, so we have stated its relation to moving averages as well as pairs trading to make our strategy more understandable. Other details include our results, assumptions as well as improvements for those who would want to pick this strategy up and subject it to more rigorous testing, something we regret not addressing in this report due to the constraints of time and resources.

INTRODUCTION

Mean reversion is a common strategy used in financial investment, which states that prices or returns will always revert to a normal pattern. Therefore, investors could make investment decisions when divergence of the market trend and its relevant mean statistics is spotted. This theory can be applied to both buying and selling, where the investor saves at the occurrence of the abnormal and profits at the upswing.

Common choice of this mean and average can be historical average of the price or return over various periods. Or it can be other aggregated statistics computed by models which take in multiple variables, such as the growth rate of a company or the average return of a sector. In our model, this average is the correlation between overnight-intraday return of a stock. We believe there exists a correlation between the overnight and intraday returns, and divergence from the correlation will yield opportunities to make profits. By optimizing the 4-Factor Model originally proposed by Zuka, we attempt to verify that such strategy will be profitable and can be applicable with improvements.

Before we go on to the strategy proper, we want to make this review readable to the average trader or investor, and thus we dedicate two stepping stones in leading up to our strategy itself.

RELATION WITH MOVING AVERAGES

Before we proceed to our main strategy, we shall use a simpler strategy for our reader to understand what and how mean reversion strategies are. We assume that the reader understands moving averages (MA). Those technical indicators are based on the concept of mean reversion, albeit only depends on the underlying asset's historical price. In the following case, we shall use the MA as the analogical stepping stone to understand our strategy better.

The theoretical use for MA is the crossover, where buy and sell signals are produced when the MA crosses over with the price; if the price is above the MA, it is deemed overpriced in relation to the calculated MA price, and thus, you should be having a bias to short the asset. Otherwise, if the price is below the MA, you should be having a bias to long the asset.

While the reader may be acquainted with the MA graphical chart that overlays the asset price chart, it is in fact, a calculated value using the price chart. This is the equation for the exponential MA:

$$EMA_t = \left(P_t - EMA_{t-1}\right) \cdot \frac{2}{n+1} + EMA_{t-1}$$

Where:

- t is the time at the particular instance of the price
- EMA_t is the exponential moving average at t
- P_t is the price of asset at t
- n is the period denoted by the user

While the equation may confuse the reader, it just shows how the more recent price affects the calculation of the exponential MA, as opposed to the simple MA.

How the MA differs to our strategy are as follows:

- We use the size, volume, volatility and momentum (WML) instead of the historical price.
- We use the logarithm value of the above 4 factors instead of the original value.
- We use the values from all the universe the asset is from (which in this case is the NYSE stocks) instead of just the underlying asset

However, in the end, in the graphical sense, it is still similar to the MA; buy and sell signals are produced when the value crosses over with the price. In our case, we just use 10 stocks that have the highest deviation to the value, 5 to long and 5 to short.

RELATION WITH PAIRS TRADING

Pairs trading is another mean-reverting strategy, betting that the prices will eventually revert to their historical trends. The difference between pairs trading and MA is that it is a market-neutral trading strategy that matches a long position with a short position in a pair of highly correlated instruments such as two stocks, currencies, commodities or options. Pairs trade is a substantially self-funding strategy, since the short sale proceeds may be used to create the long position. Pairs trading has potential to obtain profits with relatively low risk.

Highly-correlated pairs often come from the same sector because they face similar systematic risks. One example of a pair is Coca-Cola (KO) and Pepsi (PEP). In a pairs trade, we bet on the direction of the stocks relative to each other. As such, pairs trading strategies are independent of market movements and are said to be market neutral. If Coca-Cola’s stock price rises relatively to Pepsi, mean reversion theory suggest to us that the relation will eventually return to historical trends. That means under such a circumstance we would take a long position on Pepsi and a short position on Coca-Cola.

BASIC 4-FACTOR MODEL

Based on Zura's thesis, we propose a 4-factor model for overnight returns and the respective definitions. All 4 factors are constructed based on intraday price and volume data. Our 4 factors involve size, volume, volatility and momentum.

We use the model proposed by Zura Kakushadze (2015), which decomposes the overnight returns($R_{s,d}$) into 4 risk factors —- Size ($\beta_{s,d}^{1}$), Momentum ($\beta_{s,d}^{2}$), Intraday Volatility ($\beta_{s,d}^{3}$) and Volume ($\beta_{s,d}^{4}$), of which the calculations are at the footnote below[1]. $\beta_{s,d}^{0}$ which is the "market beta" is set to be always 1 to reflect that a portion of the returns of every stock is due to overall market movement. f_d^i is the specific return of risk factor $\beta_{s,d}^{i}$. $\varepsilon_{s,d}$ refers to the residuals and it is the intercept of the linear regression. *s* is the index for stock and *d* is the index for day. Thus, the formula is written as such:

$$R_{s,d} \sim \sum_{i=0}^{4} \beta_{s,d}^{(i)} f_d^{(i)} + \varepsilon_{s,d}$$

Based on Zura's research results that the overnight return is negatively correlated with the intraday return of the stocks, we buy the stocks at the open price and sell them at the close price. At the opening of every trading day, we have previous day's prices and today's opening price, from which we can calculate the forecasted overnight returns using the 4-factor model. By doing so, we are slicing the assets into the overnight part and intraday part, longing the overnight and shorting the intraday. Meanwhile, we can calculate the true overnight returns and if the true value is higher than the predicted value, we expect the stock price to mean-revert, i.e., stock price will go down. Therefore, we sell the stock during day time. The net investment amount on each stock is determined by normalization with respect to the desired dollar holdings formula.

RESULTS

Using the above strategy and our model, we ran a back-test on stock data from 29/1/2013 to 20/10/2016 using a standalone Python algorithm, and the results are as follows:

623 – 378 (62.2%) Wins Losses Win Rate	
0.294% - 1793% Average Daily Return – Cumulative Return	**$2778266** Transaction Fee
14 – 7 (10.47%) Max Consecutive Wins – Max Consecutive Losses (Max Drawdown)	**2.73** Sharpe ratio

The transaction fees used are from the Interactive Brokers platform, which includes FINRA's transaction costs. Since we used a standalone Python algorithm, there are some items missing that were not calculated such as the number of trades, which will be touched on in the Assumptions section. The transaction fees are exceptionally high because of the high frequency of trading of the underlying securities.

[1] $\beta_{s,d}^{1} = \ln(Close_{s,d-1}),\ \beta_{s,d}^{2} = \ln(\frac{Close_{s,d-1}}{Open_{s,d-1}}),\ \beta_{s,d}^{3} = \ln(\sqrt{\frac{1}{21}\sum_{k=1}^{21}(\frac{High_{s,d-k}-Low_{s,d-k}}{Close_{s,d-k}})^2}),\ \beta_{s,d}^{4} = \ln(\frac{1}{21}(\sum_{k=1}^{21} Volume_{s,d-k}))$

ASSUMPTIONS

Due to the constraints of time in building and testing out our strategy, we had the backtesting of the strategy to run on the following assumptions:

- Stocks can be divided to the nearest integer. The bigger the capital, the nearer the actual result is to our result.
- We assume that there are no commissions and a spread of $0.01 per share. The smaller the commissions and spreads are, the nearer the actual result is to our result.
- The calculations assume that the values that we use are log-normally distributed.
- We have assumed that there is no slippage in our trading, such that we can always get more favourable prices in trading. This contributes to our exceptionally high cumulative return.

IMPROVEMENTS

Some parts of our improvement plans are:

- To apply rounding to the nearest integer number of stocks for each trade,
- To add new condition for stock selection to optimize the number of stocks traded,
- To add calculations on the transaction costs and commission, and
- To consider slippage as a significance level of the test.

We will also be doing more rigorous testing by means of Monte Carlo analysis and uploading our code onto Quantopian, or use the zipline module, for testing.

Upon improvement of strategy on market timing and performance, we have been considering introducing a new risk control measure and determining the optimal frequency of trading to maximize the possible return. After the introduction of new risk control measure, our holding position formula will become:

$$H_{s,d} = -\frac{\widetilde{\varepsilon_{s,d}}}{sd(s)}\left(\frac{I}{\sum_{j=1}^{N}\left|\frac{\widetilde{\varepsilon_{j,d}}}{sd(j)}\right|}\right)$$

Mean Reversion - 4-Factor Model and Overnight Returns

Research Analysts: WENG LING, WANG ZEXIN, LI CHENGCHENG, XU CHEN, SYAKYR SURANI
