Momentum Investing and Machine Learning Applications to Factor Investing

AFM 423

University of Waterloo

Professor Tony S. Wirjanto

By:

Daniyal Kazimi (20834097)

Michael Lin (20841316)

Jeffrey Qiu (20838203)

1 Introduction

In this report, we aim to present and discuss the topic of momentum investing as a machine learning application to factor investing. In order to do this, we will provide a survey of said topic by performing a critical review of 3 published online papers in that area. Our investigation and research will be leveraged in further writing where we will aim to develop our own implementation in the topic. The following are the 3 papers we have selected for review:

1. Applying Machine Learning to Trading Strategies: Using Logistic Regression to Build Momentum-Based Trading Strategies, by Beaudan, P. & He, S.
2. Dissecting Momentum: We Need to Go Deeper, by Borisenko, D.
3. How news and its context drive risk and returns around the world, by Calomiris, C. & Mamaysky, H. (2019).

2 Applying Machine Learning to Trading Strategies: Using Logistic Regression to Build Momentum-based Trading Strategies

<u>Summary</u>

The paper discusses the approach to utilize machine learning techniques to develop trading strategies based upon the foundation of price momentum.

They begin by discussing the actual concept of momentum in the financial markets, and what challenges arise from trying to utilize momentum signals to make a profit. Some of these challenges include the issue of overfitting, where trading data is based on historical data that is too specific, and does not adjust well to current market conditions.

The datasets used by Beauden were historical price and volume data for the SPX, as well as the following seven equity indices:

- S&P Small Cap 600 Index (SML)
- S&P Mid Cap 400 Index (MID)
- TSE 100 Index (UKX)
- FTSEurofirst 300 Index (E300)

- Tokyo Stock Exchange Price Index (TPX)
- Dow Jones Industrial Average Index (INDU)
- Dow Jones Transportation Average Index (TRAN)

The methodology involves comparing two base cases, one in which the investment decision is made to be on a specific date, regardless of conditions, while the other is a dual-momentum strategy, balanced monthly. These are compared to the machine learning approach, which utilizes a logistic regression analysis to predict the probability of each asset experiencing positive momentum. The testing was also split into two further subdivisions, 'long-only' and 'long-short', where the former would only buy assets with positive momentum signals, while the latter would sell short assets on negative momentum signals in addition.

The results showed that over the *full* testing period (1964-2018), cubic logistic regression yielded a higher annual of 8.6% and 10% for the long-only and long-short strategies respectively, as opposed to the dual momentum strategy yielding 4.87% and 4.07%.

Strengths

The data is applied to a wide historical period, meaning that it is likely relevant as long as market conditions are not massively different over a wide period. Furthermore, the presentation of the results are well done, with the full period being split into multiple smaller ones in which the returns of the various methods can be compared, as well as including other useful metrics such as volatility, Sharpe ratio, and more.

Weaknesses

Only momentum strategy was focused on, it would have been interesting to have a few other strategies included for comparison. The data is solely based on historical price data, where the effects of other variables being taken into account may have an impact, such as fluctuations in commodities. The issue of overfitting remains, where uncommon market conditions may result in the model not being a good fit

3 Dissecting Momentum: We Need to Go Deeper

Summary

This paper dives into the concepts and theory behind momentum. In finance, momentum refers to the phenomenon of cross-sectional predictability of returns by past price data. Over the years, many ways to measure movement have been proposed. This includes different lookback horizons and various factors explaining or predicting its risk and return. Borisenko aims to construct a broad set of price characteristics over various time horizons and investigate their predictive power in a deep learning framework.

Strengths

This paper does an excellent job measuring price momentum in a broad manner. Compared to other machine learning asset pricing literature Borisenko referenced (Gu et al. (2018), Messmer (2017)), his construction of a broad set of price characteristics over various time horizons adds meaningful value. Borisenko makes contributions not just in an empirical nature, but also methodologically and practically as well.

Relevant to course content, in this paper Borisenko develops a deep learning model and expands on it with great detail. He contends that investment strategies built on its predictions can "actively exploit the non-linearities and interaction effects, generating high and statistically significant returns with a robust risk profile and their performance virtually uncorrelated with the established risk factors involving momentum." Not only is this an insightful model to see created and learn about, but Borisenko further discusses the potential shortcoming he observed that the model is at risk of overfitting the training data. Overfitting training data is a common concern with learning models, in which we've observed to a degree. To account for this, Borisenko employs a "dropout" technique (pioneered by Srivastava et al. (2014)) and early stopping techniques. This self-acknowledgement and solution for potential concerns about his model helps give further credibility to the paper.

The paper's major strength and the one which experts in the field would most appreciate is the level of detail and rigour. Borisenko clearly demonstrates that he himself is an expert in the field through sound conceptual explanations and the development of his own model. By using established literature as a foundation for his work in this paper, he gives readers relevant background from which his own findings and analysis can be derived. On the technical side, the paper does well in describing the model and how it was trained with significant depth. Lastly, Borisenko makes a case for adoption of automated hyperparameter optimization techniques as an important component of disciplined research in financial machine learning, a highly advanced topic.

Weaknesses

Although it is largely inherent with the subject matter, the article is incredibly technical and complex. Many readers may not understand the concepts presented and they require strong background and knowledge to be able to derive valuable meaning. While not necessarily a "weakness" of the paper, as it may be apt for its target audience, more layman terminology and a high level overview of the paper would be a great asset for those seeking a less in-depth look on the topic.

As a note to the reader, Borisenko addresses a shortcoming in wording. As he says, he uses "classification instead of regression and the phrases like 'predicted probability of return being above the cross-sectional median increases in X' does not roll off the tongue, to say the least, though it means that 'the expected return in X'". To workaround this, Borisenko himself suggests defining something such as '*expected return*' to mean the expected return in terms of the predicted probabilities. Defining this at the beginning of the paper and using it throughout the text would help make the wording less cumbersome. I agree with his opinion and while not a substantial issue, the paper would become easier to read through and more digestible.

4 How news and its context drive risk and returns around the world

Summary

The paper discusses how news can be leveraged for asset pricing (specifically market returns) and key elements that should be considered. The paper makes use of concepts learned through the course such as penalized regressions in creating their predictive model. The nine key questions addressed in the paper are:

- How to best measure news using words? Emphasis placed on methods without priori positions to avoid unnecessarily restricting their analysis.
- What aspects of words should be focused on? Aspects analyzed in depth are sentiment, frequency, topic/context and entropy (unusualness).
- What patterns link the various measures/aspects of words? Makes use of elastic net regression for model selection based on 1998-2015 sample.
- How should words be classified based on topic? Focused on the Louvian method (mutually exclusive topics) due to simplicity and similar outcomes.

- How does word flow influence risk? Addresses the issue of positive news (higher expected return) corresponding with lower risk and vice-versa.
- How should risk be measured? Paper uses one-year ahead drawdown (longer-term) in addition to monthly volatility (shorter-term).
- What impact does news have across countries? The paper groups countries into either developed or emerging markets.
- What sources of news should be included in the dataset? Thomson Reuters, an international English language news source, is used as the source of news articles.
- What time horizon should be used for predictions of risk and return? The paper finds that aggregating news over longer time periods (daily vs weekly vs monthly etc.) allows for longer term predictions.

Strengths

The sample period is expansive (1998 to 2015), accounting for various economic events such as the Dotcom and Great Recessions as well as periods of economic growth and recovery. Within the sample, articles that are related but not exclusively focused on the topic areas determined by their algorithm are kept. These factors allow for a more comprehensive sample and testing sets which are used to create their model.

The paper is structured in a step by step manner ensuring the methodology can be followed easily. The authors limit the use of overly technical information in the main paper (included in the appendix), to avoid breaking the flow of their paper. Calomiris and Mamaysky also reference a variety of sources and alternatives from a variety of authors in justifying their assumptions and methodology. These indicate much thought was put into structuring and supporting their analysis in a way that is readily more understandable and accessible to less experienced readers.

Weaknesses

While practical for the purposes of text analysis, the restriction to one English language source for data, could incorporate significant biases. These could be inherent to the reliance on an outsider as a source of news (articles for non english speaking countries) and the influence of editorial policies at Reuters.

What is missing

While intuitive from a macroeconomic perspective, limiting the comparisons to emerging vs developed markets may restrict applicability of the model. An example of this is natural disasters which can be very localized within a certain geographic area. Some countries also form economic blocs like the EU whereby the economies within a certain region are more closely linked. An example would be the Czech republic (categorized as emerging) whereby news from a nearby developed country (Ex: Germany) would be much more relevant than news from emerging markets across the globe (Ex: China). Emerging economies are also a very diverse category with rapidly growing economies (Ex: India) being grouped together with stagnating ones (Ex: Argentina), and already fairly developed markets (Ex: Israel). Such observations indicate that the data could be recategorized based on geographic regions to improve predictive capabilities or at least a viable alternative.

5 Connecting the Dots

All the papers talk about the concept of economic momentum in regards to assets, and they also talk about how price momentum is not necessarily an objective measure. While the articles published by SSRN focus more on the quantitative side of price momentum, mainly focusing on historical price data for groups of assets such as the SPX, the article published by the Journal of Financial Economics takes a deeper look at the qualitative factors that go behind price momentum, such as news and current events, which is harder to account when looking at the numbers alone. However, while the specifics on how price momentum works may differ slightly in the articles, it is agreed that price momentum is a real-world observable phenomenon, one that can be used to formulate an investing strategy.

6 Plan

In order to achieve the objective of gaining hands-on experience within the topic area, our plan for the second portion of the report is as follows: To develop a momentum-based investment learning model. Our project aims to determine the most effective time horizon among 2-day, 7-day, and 30-day momentum signals to generate profitable returns, if any. Furthermore, we will look at certain 'unusual' periods in which market conditions were abnormal, and see how our model performs in such conditions.

To achieve this, we will train a learning algorithm for the model using historical market price data for financial equities markets such as the S&P 500. We will also compare draw comparison against a theoretical investor with a more traditional strategy, such as a dollar-cost-average strategy, to determine if the model outperforms a traditional investor.

We will leverage course concepts such as linear regression analysis, graphic visualization, and learning algorithms.

References

Beaudan, P. & He, S. (January 30, 2019). Applying Machine Learning to Trading Strategies: Using Logistic Regression to Build Momentum-Based Trading Strategies. Available at SSRN: https://ssrn.com/abstract=3325656 or http://dx.doi.org/10.2139/ssrn.3325656

Borisenko, D. (July 23, 2019). Dissecting Momentum: We Need to Go Deeper. Available at SSRN: https://ssrn.com/abstract=3424793 or http://dx.doi.org/10.2139/ssrn.3424793

Calomiris, C. & Mamaysky, H. (2019). How news and its context drive risk and returns around the world. Journal of Financial Economics, 133(2), 299–336. Available at https://doi.org/10.1016/j.jfineco.2018.11.009

Wirjanto, T. S. (2023). Topics in Financial Econometrics.