Shapley Attribution in Machine Learning Trading Models

Exploring Factor Style Attribution for Better Insights into Model Behavior

Affor Analytics

April 2024

Introduction

The rise of complex machine learning models has introduced a challenge: understanding how these models arrive at their predictions. This is because the algorithms behind these models have grown more sophisticated, making the models' predictions hard to interpret. That's why machine learning models are often called "black boxes" – their inner workings are hidden.

However, understanding how specific features and datasets influence a model is essential, also when explaining and optimizing trading strategies. If a particular dataset consistently produces inaccurate results, we can either remove it, adjust it to address the inaccuracies, or supplement it with new data.

For example, consider a strategy based on factor selection and timing. A simple 50/50 blend of traditional value and growth portfolios makes it easy to attribute returns: 50% to value factor style and 50% to growth factor style.

However, when using more advanced strategies driven by machine learning algorithms, uncovering these attributions becomes more difficult. These models showcase complex relationships and non-linear patterns among features, making it challenging to separate the contributions of individual factors.

Traditional methods for calculating attribution, like "one-at-a-time"[1] and "leave-one-out"[2], suffer from biases when inputs are correlated or interact with each other.

Therefore, in this longread, we explore how Shapley Attributions can be used to address the "black-box problem" without the typical biases from traditional methods. The Shapley Attributions model makes it easier to assess performance, and thus improve quantitative trading strategies that use machine-learning models.

1. The "One-at-a-time" method isolates specific inputs to make predictions based solely on each individual input in succession.
2. In the "Leave-one-out" approach, a specific input is excluded, and the strategy is applied using all other inputs, allowing for an assessment of performance when each input is omitted iteratively.

The Shapley Attribution Model and its benefits

Shapley Attribution is an extension of Shapley Values, originating from cooperative game theory. Shapley Values measure each player's contribution to the overall game [1]. Similarly, in machine learning models, Shapley Attribution evaluates the contribution of individual features to predictions, relating these values to model accuracy.

This approach addresses the limitations of traditional attribution methods by offering unbiased assessments of specific inputs to accuracy. It evaluates the contribution of each input, including their interactions, and how it affects the overall accuracy of a model [2]. Through this method, we can evaluate the attribution of different inputs and their interactions, offering a comprehensive understanding of their impact on the model's accuracy.

Moreover, Shapley Attribution not only provides unbiased estimates of how each input contributes to the model's accuracy but also ensures that when you add up these contributions, you get the total accuracy of the model. This means that by summing up the Shapley Attributions, you can directly understand how much each input impacts the model's overall performance. In contrast, traditional methods only offer relative attributions, making it harder to gauge the true significance of each input.

When evaluating a model, we look at how well its resulting trading strategy performs. We can use the Shapley Attribution method directly on the strategy's return. Because attributions add up, we can organize inputs as we like. For example, grouping features by factor style helps us understand how each factor style affects returns. This insight is crucial for improving the model and informing investors.

An example of the Shapley Attribution model in action

To demonstrate the effectiveness of the Shapley Attribution model in understanding model performance, we use a simple out-of-the-box XGBoost[3] model built on our categorized feature databank[4].

Our feature databank includes categories such as momentum, value, growth, sentiment, technical, and quality, contributing to the complexity of interpreting performance due to their non-linear relationships and interactions.

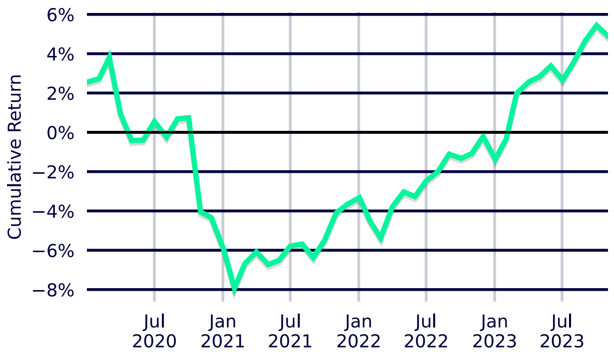

Figure 1 - Backtest results for the out-of-the-box model. These are provided purely for demonstrative purposes to showcase Shapley Attribution and do not reflect real-world accuracy.

January 2021 Performance Analysis

Considering backtest results from Figure 1, investors tend to focus on the bad months. We therefore analyze January 2021 as an example[5]. During this time, our strategy had a return of -2%.

January 2021 marked a month with heightened market volatility attributed to various economic, political, and global health-related uncertainties, including the emergence of the COVID-19 vaccine. High

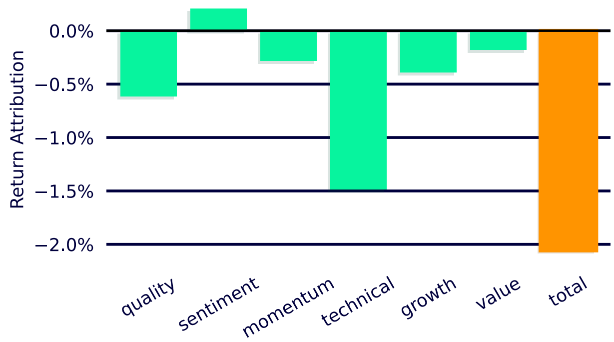

Figure 2 - Shapley attribution scores to return for the January 2021.

volatility can lead to unpredictable market movements, making it difficult for predictive models to forecast accurately.

In Figure 2, we observe the Shapley attribution scores for our example, grouped by different categories. Notably, the technical category[6] stands out with the largest negative attribution. This Shapley score of -1.49% indicates a decrease in performance attributed to the use of technical features by the model. Our attention to the technical group is justified as it plays a significant role in the model's return, alongside quality.

The Shapley attributions enable us to connect the performance of underlying features, specifically technicals, with macro events. It's likely that our technical features underperformed in January 2021 due to sudden market shifts triggered by external events, such as the emergence of the COVID-19 vaccine and the resulting economic uncertainties.

Factors influencing performance

Rapid change in market sentiment and increased volatility

Technical analysis is largely based on historical data and patterns. When new, unexpected events occur, market sentiment can shift rapidly, making historical patterns less predictive of future movements.

The announcement of the vaccines led to increased market volatility [3]. Technical indicators can sometimes

3. It's important to distinguish between feature attribution and feature importance in tree-based models. Feature importance refers to the overall influence a feature has on a prediction, quantified by its total weight. However, this is nondirectional; a high feature importance doesn't necessarily indicate that the feature made accurate predictions.
4. Affor Analytics databank consists of 100+ features using fundamental, sentiment, and technical data.
5. November 2021 is also a bad month, but Shapley Attributions are inconclusive.
6. Examples of technical features are realized volatility, skewness, and turnover.

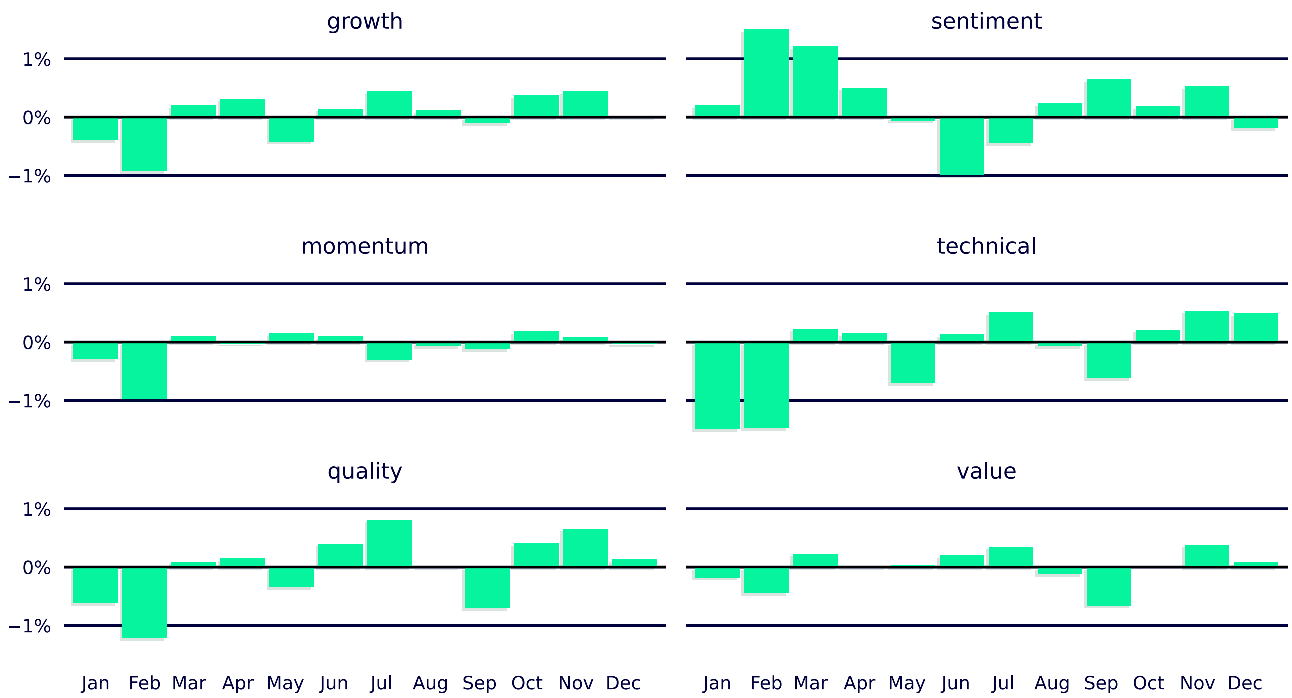

Figure 3 - Shapley attribution values to return for factor styles over 2021.

provide misleading signals in highly volatile markets because the underlying assumptions of stable trends and patterns become less applicable.

Supply Chain and Labor Market Disruptions

The economic data released in January 2021 tells us that there were supply chain and labor market disruptions [4]. Such macroeconomic factors can have a significant impact on company performances and, consequently, stock prices, which may not be immediately or adequately reflected through technical indicators.

Performance attribution of style over time

It can be valuable for both model interpretations and improvements to plot Shapley attribution over time. Figure 3 shows that January and February were by far the worst month for the technical group. Except for a few major drawdowns the attribution of the technical feature in 2021 is rather small. This might indicate that including technical features adds much risk and is therefore not a good addition to the model used.

Other categories never show such a big negative Shapley score. However, most styles do still show some months with big drops in attribution to total return. Both quality and momentum have some major negative scores. These cases can be used to investigate a recurring pattern, which could then be used to improve the model.

Shapley Attribution for better risk management and trading strategies

Machine learning and complex algorithms are here to stay. The Shapley attribution is a powerful tool for understanding how different factors contribute to the performance of quantitative trading models. It offers a detailed and unbiased assessment of each input's impact, considering the complex interactions between features. This not only improves transparency but also provides insights for refining trading strategies.

Shapley attribution goes beyond analyzing returns to assess various risk metrics, aiding in more nuanced risk management. It can pinpoint specific features associated with high volatility, helping traders make informed decisions.

Moreover, Shapley attribution isn't limited to factor styles; it can be applied to any input, including individual securities or features. It enables the detection of seasonal patterns, which can inform seasonal-based input selection strategies. This allows for thorough evaluation even in scenarios involving multiple models or during portfolio construction.

In essence, Shapley attribution offers a practical framework for enhancing quantitative trading strategies by providing clear insights into performance attribution.

References

1. Lundberg, S. M., & Lee, S.-I. (2017, December). A Unified Approach to Interpreting Model Predictions. NIPS conference, Advances in Neural Information Processing Systems volume 30.
2. Sundararajan, M., & Najmi, A. (2020, November). The many Shapley values for model explanation. In International conference on machine learning (pp. 9269-9278). PMLR.
3. Investment Communications Team (2021, February). Monthly markets review - January 2021. Schroders.
4. Ulate, M., Vasquez, J. P., & Zarate, R. D. (2023, February). Labor Market Effects of Global Supply Chain Disruptions. Federal Reserve Bank of San Francisco Working Paper Series

Disclaimer

This report is prepared and circulated for educational and informational purposes only and is neither an offer to sell nor a solicitation to invest. Past performance is not indicative of future results. The value of investments and any income generated may go down as well as up and is not guaranteed. The report does not constitute any recommendation, investment proposal, offer to provide a service, solicitation to buy or sell any security or other investment product. The information may be subject to restrictions imposed by law in some jurisdictions. This information is not meant for visitors within a jurisdiction in which such an offer or solicitation is against the law, or for a person to whom it is unlawful to make such an offer or solicitation. Affor Analytics accepts no liability for infringement of such restrictions. The recipient shall not distribute, forward or publish the information. No rights may be derived from the provided information, data, and calculations. The value of the products is (among others) subordinated to the developments on financial markets and, if applicable, other markets. As a recipient of this report, you acknowledge the disclaimers contained within this report, you acknowledge that the data and contents of this report are provided on an 'as is' basis and you assume the entire risk of any use you make or action you take based on this report ore data. Affor Analytics and/or the authors of this report are in no event responsible and thus liable for any direct or indirect damages, including, without limitation, any lost profits, lost savings, or other accidental or consequential damages that follow from your use of the findings in this report.

Authors

Jasper Kousen
Head of Data & Technology
j.kousen@afforanalytics.com

Koen Ripping
Director
k.ripping@afforanalytics.com

Isaac Braam
Junior Data Scientist
i.braam@afforanalytics.com

Contact

Affor Analytics

Stationsplein 45, A4.004
3013 AK, Rotterdam
Netherlands

ir@afforanalytics.com
+31 10 766 00 50