

Explainable Machine Learning for Stock Return Prediction- Taking Apple Data as an Example

Yuhan Wei

*Sidney Institute of Language and Commerce, Shanghai University, Shanghai, China
24124258@shu.edu.cn*

Abstract. Stock return prediction is a fundamental but difficult issue in financial research because of the noise and dynamics of market data. Although recent research has improved the ability of existing machine learning model prediction, these methods are difficult to interpret and lack practical application value due to their "black box" nature. This paper constructs an explainable machine learning framework to achieve high-accuracy predictions of short-term stock excess returns and strive for greater interpretability in economics; Compared with other complex models. Taking the daily trading data of Apple Inc. (AAPL) as a single-stock example, this paper constructs a set of economically meaningful features based on historical returns, volatility and trading volume. A linear regression model is used as the central prediction tool, and interpretability can be built into this specific form of a linear model. Out-of-sample forecasts use the rolling window method, and prediction performance is compared with a naive historical mean benchmark. Based on experiments, the current basis of explanation using an explainable linear model has been shown to be relatively stable in baseline prediction. According to the coefficient analysis, in addition to conforming to some basic Finance theory patterns. A return pattern that exists for one day appears. A Volatility-Return relationship also holds currently. This paper shows that interpretive models, in combination with carefully selected features, achieve satisfactory prediction performance; at the same time, it is also found through the analysis of economic implications that such an approach helps explore the essence of changes in stock returns.

Keywords: Explainable machine learning, stock return prediction, Apple Inc.

1. Introduction

Stock return prediction has always been a central issue in financial economics that is closely related to asset pricing, portfolio construction, risk control and other fields. Although there has been more than a decade of research on this issue, it is still difficult to achieve an accurate short-term return forecast due to market noise, the fast-paced update of information, and the complexity of investors' psychology. The traditional econometric model, such as the linear regression and factor method, has strong theoretical interpretability; But in general, it is not easy to explore all details deeply hidden in high-frequency financial data.

In recent years, the application progress of machine learning has brought more flexible model construction methods that can describe non-linear relationship characteristics or realise complicated

inter-variable relationships in finance. Empirical research shows that the classical benchmark usually has poor performance in out-of-sample forecast tests for machine learning and deep learning models. There is no doubt this problem will remain invisible to anyone else. Many machine learning models are opaque in their operation, and it is hard to know how they make predictions, or if the information they extract has economic significance. Lack of interpretability makes them less applicable to practical application in financial decision-making, which requires trust, accountability and theory support at the same time.

In response to this, explainable Artificial Intelligence (AI) has emerged with an aim to make the behaviour of the prediction model more transparent and understandable for people. The explanation ability of the application in finance is essential because model output results need to be explained regularly during investment decisions and comply with regulatory requirements or evaluate risks. Although some studies have put forward post hoc explanation methods or interpretable deep learning structures, explainability is generally regarded as an auxiliary link rather than an indispensable step in model building.

To summarise this study, using an explainable machine learning approach to build a model for predicting stock return data is expected to add new content to current research. In comparison with the emphasis on high-order complex model building, this paper focuses primarily on promoting its practicality for prediction engineering, along with simplicity, economy, etc. in the entire flowchart of predictions. Economically motivated features are built according to the historical price movement and trading activity rules, and a linear regression model is adopted to achieve transparent coefficient explanation. Taking Apple Inc. as a typical large-cap stock, this paper examines its predictive performance in an actual time series through rolling window out-of-sample forecasting.

The empirical study shows that a simple, highly interpretable model performs better than the naive baseline and is in line with existing financial theory. Through the combination of explainability embedded in feature Design, model Specification and Result Interpretation, this paper explores the significance of transparent machine Learning approaches in financial forecasting and offers an operational solution to the black box Prediction problem.

2. Literature review

2.1. Stock return prediction methods

Stock return prediction has always been a central issue in financial research; early studies mostly used traditional econometric and statistical methods. The traditional linear regression model, AR model and the factor approach are all commonly used methods that try to decompose an individual stock's return into several predetermined factors. These methods are appreciated for their theoretical foundation and interpretability, but they may struggle to perform well when there is an actual nonlinearity and complex market environment [1].

In response to the above problems, researchers have gradually focused on the Data-Driven method and relaxed the strict parametric assumption. In recent times, some research have started to integrate other kinds of information into consideration when building models; that is to say, they incorporate technical indicators and higher-order return data for better prediction capabilities. Traditional models are unable to fully describe the nonlinearity and non-stationarity of financial time series, so more flexible prediction systems have been introduced [2]. Therefore, literature has gradually moved towards using machine learning techniques for stock return predictions.

2.2. Machine learning in financial applications

Due to the ability of machine learning (ML) and deep learning algorithms to represent complicated nonlinear relations, they have received increasing amounts of focus in research on financial prediction issues. Various types of model tree, neural network, and ensemble learning methods have been widely used for predicting stock returns or price changes; They usually show excellent out-of-sample performance relative to traditional benchmark indexes [2,3].

Recently, some studies have broadened their research areas by combining machine learning technology with other data sources; that is to say, they apply it to news text or financial report information. Large language models (LLM) and text-based deep learning architectures can obtain informational Signals from Unstructured Data, which helps improve prediction Performance [4,5]. Moreover, some new architecture ideas of graph type and contrastive learning have also been proposed to mine these two types of dependencies more fully across time series [3].

Although there have been some achievements in the application of machine learning to financial models, a common criticism is that it is a "black box". Although the prediction accuracy has increased, due to its lack of transparency, it is not easy to use in actual application; In terms of transparency, interpretability and confidence are very necessary factors for decision making and risk management.

2.3. Explainability in financial prediction models

As a type of explanation-based AI, XAI has attracted more and more attention in financial forecasting because its lack of transparency makes it difficult for many complex machine-learning models to be applied effectively at work, and there is no real improvement in application; opacity also hinders decision-making and risk control [6,7]. Recently, some research has attempted to improve the transparency of models via feature importance analysis, attention mechanism, and interpretable deep learning architecture [6,8]. Keyword-guided and factor-based frameworks try to strike a balance between the rationality of data-driven conclusions and investors' financial instincts [9-11].

Although some progress has been made in addressing this issue so far, most existing research still views explainability as an additional step following the construction of prediction models; that is, it does not integrate interpretability throughout the entire modelling process. Few studies have systematically incorporated explainability into the entire process of feature construction, model selection and results interpretation.

To fill this void, the present study has developed an integrated explainable model framework that incorporates interpretability into the entire prediction process. Specifically, economically significant features are built based on price dynamics and trading activity. A linear regression model that is inherently interpretable is selected; And the model output is explained using coefficient interpretation. Integrating explainability into three stages of feature design, model specification, and empirical interpretation, it is expected that the prediction result will be more transparent and have economic meaning; At the same time, it also provides a coherent basis for the methodological Design and empirical Analysis in the subsequent section.

3. Methodology and research design

3.1. Data description

Using the daily stock market data of AAPL as a single-stock case, this paper demonstrates the proposed explainable machine learning framework for stock return prediction (Figure 1). The dataset contains standard daily trading information, including the opening price, highest price, lowest price, closing price, and trading volume. Selecting only one highly liquid stock makes it easier to focus on the transparency and interpretability of the methods in this paper; at the same time, people can avoid dealing with the issue of cross-sectional heterogeneity.



Figure 1. AAPL daily closing price

As shown in Table 1, the choice of AAPL is driven by its large market capitalisation, steady trading volume and broad application in empirical financial research; thus, it serves as a typical example to illustrate the methods. All data is arranged in a sequential manner over time, and general cleaning processes need to be applied at the beginning to maintain data integrity. Days with missing data are deleted, and the dataset is arranged according to the information available to investors at different points in time to avoid forward-looking.

Table 1. Summary statistics of raw variables

	Mean	Variance	Std.Dev.	Min	Max
High Price	235.9842	719.2936	26.81965	190.335	288.62
Low Price	230.8684	743.6851	27.27059	169.2101	283.3
Trading Volume	5.39e+07	5.37e+14	2.32e+07	1.79e+07	1.84e+08

3.2. Variable construction and feature engineering

The primary prediction objective of this study is the one-period-ahead stock return. The daily return is calculated based on the logarithmic return formula, that is, the difference in the log of successive closing prices. Log returns are often used in financial Research to add and approximately normally distributed.

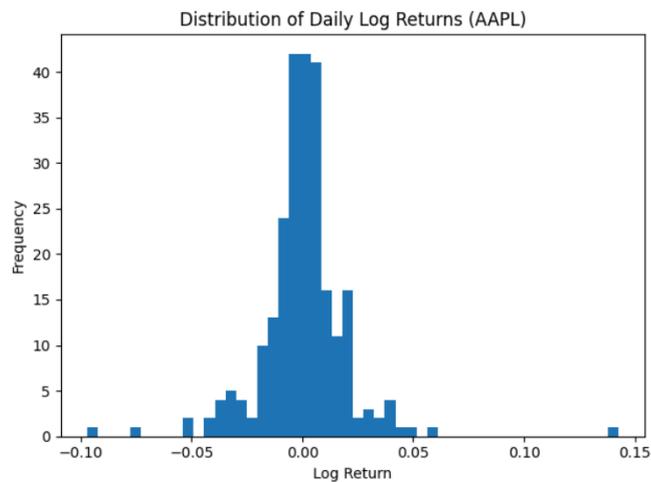


Figure 2. Distribution of daily log returns of AAPL

To build explanatory variables, this study will focus on two aspects: one is to select features with high informativeness; the other is to choose economically interpretable ones. Lagged returns are introduced for two reasons: one is to capture short-term momentum, and the other is to take advantage of the literature's documented effect of price reversal. The second type of simple technical indicators based on historical data price information includes moving averages and the degree of change in returns over a specific rolling window. These indices summarise the trend of recent prices and the degree of market uncertainty in a simple way (Figure 2).

Trading Volume is introduced to represent the degree of market Activity and investors' participation. Volume-type variables generally reflect the flow of information in the market and the extent to which investors participate; it is possible that these factors will have some influence on short-term return levels. All feature construction uses only historical and current Data, so the model is compatible with actual applications for prediction. Given that interpretability of variables has been prioritised in the literature on explainable models, the feature engineering process should maintain this alignment.

3.3. Model specification

Given that the goal is to achieve a balance between interpretation and economic transparency, this study uses the linear regression model as the core prediction framework for stock return forecasting. Linear regression is a natural reference model in financial research because it can directly map the explanation variables to the expected return with interpretability. Based on the estimated coefficient can roughly determine in what direction, with approximately how large a scale Does each predictor influence the value of this model; help identify their characteristics.

In other words, this type of prediction task falls under supervised learning, where the target variable for training is the one-period-ahead stock return; The explanatory variables include various features that interpretable construction from historical Price Dynamics and Trading Activity patterns. Lagged returns, Volatility-related factors, and Volume-based factors have been widely used in traditional finance.

The selection of the linear model has been premeditated. Rather than seeking the maximum predictability, this paper is more focused on method simplification and economic interpretability. Linear regression provides an interpretable coefficient form to explain it, is easy to compare with traditional financial models, and reduces the risk of overfitting in a small sample size, single asset

scenario. In order to control the complexity of the model and avoid meaningless predictions of relationships in the analysis that cannot be linked to observable market behaviour, this step should be taken.

To sum up, the linear regression specification falls under that of the explainable machine learning methods applied in this paper because its interpretability is directly embeddable in the model structure. This Design selection provides a reference for the following out-of-sample Test and explanation analysis so that both factors can be considered at the same time: prediction performance and model interpretability; It is not to treat these two goals as contradictory.

3.4. Model training and evaluation strategy

To make an objective assessment of the prediction effect, this paper takes time-series-aware data division strategy based on the order of observation chronologically. The daily data of March 2025 to October 2025 are used as the training set (168 cases), and the observation period from November 2025 to January 2026 is selected as the out-of-sample test set (76 cases). The time-based partition can avoid look-ahead bias and reflect the information structure that real-world investors face.

Out-of-sample forecasts are produced using the rolling window method and setting a fixed window length of 168 trading days. For each forecast origin, reestimate the model with the latest 168 observations and then use it to predict the next period's return. Iterate this procedure during the test period to obtain a series of actual out-of-sample prediction results.

Forecast accuracy is assessed through mean squared error (MSE) and mean absolute error (MAE) for the test set. All features and rolling-window updates depend only on the information available up to each prediction point, and no future information is used during model training or evaluation.

3.5. Explainability methods

Explainability is at the centre of the provided structure. In addition to assessing the prediction's accuracy, it is necessary to determine which feature components of an individual are related to predicting the model. For linear models, the estimated coefficients directly achieve interpretability; That is, it can be determined that an increase or decrease in each explanatory variable will cause an increase or decrease in the value of the dependent variable, and this change magnitude can also be estimated.

The following table summarises important attributes according to their contributions when making predictions with this model and ranks them by significance. The obtained importance scores represent the degree to which each feature can decrease the prediction error of the model in its response capability. These measures can give people some visual hints about which dimensions of price movement and market situation are more closely related to the return forecast.

To combine inherently interpretable models with explicit explanation methods in an attempt to incorporate interpretability into the model building process and not handle it separately after completion. To help people understand economically the meaning of this research results in terms of theory and provide a premise for subsequently analyzing predictive capacity and giving an explanation accordingly.

4. Empirical results

4.1. Prediction performance

This part presents the out-of-sample predictive performance of the proposed explainable linear regression framework for stock return prediction. To provide a reference, the model forecast is compared with a naive historical mean strategy that predicts future returns based on the average of past data. The prediction accuracy is assessed in terms of out-of-sample test period mean square error (MSE) and mean absolute deviation (MAE).

Table 2. Historical mean of prediction performance

Variable	Obs	Mean	Std. Dev.	Min	Max
sqerr_mean	76	0.0000998	0.0001132	1.59e-08	0.0006355
abserr_mean	76	0.0085046	0.0052717	0.0001262	0.0252092

Table 3. Linear regression of prediction performance

Variable	Obs	Mean	Std. Dev.	Min	Max
sqerr_linear	76	9.39e-06	0.0000113	8.04e-11	0.0000528
abserr_lin~r	76	0.0025229	0.0017517	8.97e-06	0.0072674

Prediction error for the historical mean benchmark is listed in Table 2. As expected, the naive strategy has relatively large forecast errors and is unable to capture the short-term fluctuations in stock returns. The current situation serves as a basis for comparison; it has not been confirmed whether the improvements in interpretability and predictive power mentioned above can actually achieve this purpose in practice.

Table 3 presents the out-of-sample performance of the linear regression model estimated with the rolling window method described above. Compared to the historical mean benchmark, the linear model has achieved much lower MSE and MAE, which indicates that it has improved forecast accuracy significantly. Based on the above results, it can be seen that a parsimonious and fully interpretable model still obtains useful prediction signals from historical Price and Volume data.

Above average error levels, the linear regression model exhibits stable performance over the entire evaluation period. The rolling estimate method causes the model's parameters to adjust in response to changes in the market situation; this method has some degree of explanation. The consistency of the out-of-sample results shows that the linear specification has good robustness in a short-horizon return prediction task.

In summary, the experimental evidence shows that the proposed explainable linear regression framework has achieved better prediction performance than a simple benchmark method; At the same time, this framework also provides full transparency for the model's structure and output content. Based on these results, this paper can conclude that meaningful predictive performance and economic interpretability do not have to be contradictory, when appropriate feature design and evaluation are conducted continuously over time series data.

4.2. Explainability results

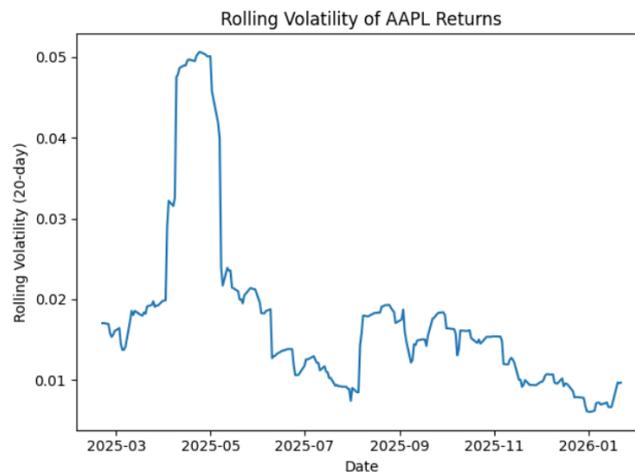


Figure 3. Rolling volatility of AAPL returns

Exploratory data analysis can provide an explanation for why people need to estimate the coefficients in a linear regression model; This is done by linking financial theory to interpret the meaning of the estimated coefficient values. Interpretability is embedded directly in the model structure, so coefficient estimates can offer clear insight into the contributions of historical prices and trading information to short-horizon return predictions.

The coefficient of the one-period lagged return is positive and has predictive power for next-period returns. This finding is consistent with the short-term momentum effect well established in the finance literature that stock returns tend to continue in a particular direction over a brief period. During daily trading, this kind of behavior is prone to gradually diffuse the information or take some time for investors to respond after receiving new information.

Volatility-related variables can help explain the pattern of returns. The Volatility of the entire market will increase when there is greater ambiguity in people's hearts and faster flow of information; these changes will directly impact individual investor estimates of potential profits from investment activities. Figure 3 shows that the fluctuation pattern of AAPL's return volatility at each time point is considerable, so in order to make accurate predictions based on this model, a Volatility-based approach has been selected.

The explanation factor of the volume shows a weak effect. The reason is that Apple Inc.'s stock has a large size and high liquidity, with an average daily trading volume; therefore, small fluctuations in the trading volume of this single day are unlikely to carry additional useful information beyond what investors have already priced in. Therefore, this result needs to be interpreted on a per-stock basis and does not apply to less liquid or small-cap stocks, where trading volume is more directly reflected in the formation of prices.

In summary, based on the explanation results, the linear regression model is able to obtain economically meaningful relationships and aligns with established financial intuition. Based on empirical research and the integration of financial theory and market characteristics, this paper has demonstrated that interpretable models help people uncover the laws behind stock returns.

5. Conclusion

This paper presents an explainable machine learning framework to explore the factors affecting short-term changes in stock returns; at the same time, it is necessary to enhance transparency and improve economic interpretation. Taking daily data of Apple Inc. as an example, this study shows that the linear regression model combined with well-designed and economically motivated features can achieve good out-of-sample prediction effect and still maintain the full transparency of the model.

Experiments show that among all compared objects, the proposed framework had consistently better performance than the naive historical mean. Explainability analysis also reveals the corresponding predicted relationship in line with established financial Theory. A positive lagged return has a significant impact and suggests that the market demonstrates a short-term momentum effect; At the same time, it was found that volatility-related factors also influence people's adjustment of their return expectations in response to changes in market uncertainty. On the other hand, the relatively weak explanation ability of trading volume reflects that these large-cap Stocks, such as Apple, have higher liquidity; therefore, when interpreting empirical results, people need to consider the influence of the market environment.

In general, prediction performance does not have to compete with explanation ability. Add an explanation function to the model construction to provide people not only with a practical prediction tool but also economic insights. In future research, the framework can be extended to cover more assets, or it can be applied in other market situations and at different times for forecasts; accordingly, its scalability will help people further examine how explainable techniques can be integrated into financial analysis.

References

- [1] Anonymous Authors (2025) Significance of Predictors: Revisiting Stock Return Predictions Using Explainable AI. *Annals of Operations Research*.
- [2] Chen, X. (2024) Machine Learning Based Stock Return Prediction and Portfolio Research. *Operations Research and Fuzziology*, 14, 1–11.
- [3] Duan, Y., Wang, W. and Li, J. (2025) FactorGCL: A Hypergraph-Based Factor Model with Temporal Residual Contrastive Learning for Stock Returns Prediction.
- [4] Guo, T. and Hauptmann, E. (2024) Fine-Tuning Large Language Models for Stock Return Prediction Using Newsflow. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, 1028–1045. Association for Computational Linguistics.
- [5] Kim, J. and Park, S. (2025) IKNet: Interpretable Stock Price Prediction via Keyword-Guided Integration of News and Technical Indicators.
- [6] Kumar, P., Hota, L., Tikkiwal, V.A. and Anand, A. (2024) Analysing Forecasting of Stock Prices: An Explainable AI Approach. *Procedia Computer Science*, 2009–2016.
- [7] Saberironaghi, M., Ren, J. and Saberironaghi, A. (2025) Stock Market Prediction Using Machine Learning and Deep Learning Techniques: A Review. *Applied Mathematics*, 5, Article 76.
- [8] ScienceDirect Authors (2024) Explainable Deep Learning Model for Stock Price Forecasting Using Textual Analysis. *Expert Systems with Applications*, 249, Article 123740.
- [9] ScienceDirect Authors (2024) An Explainable Deep Learning Approach for Stock Market Trend Prediction. *Heliyon*, 10, Article e40095.
- [10] Wang, M., Izumi, K. and Sakaji, H. (2024) LLMFactor: Extracting Profitable Factors Through Prompts for Explainable Stock Movement Prediction.
- [11] Yu, X., Chen, Z., Ling, Y., Dong, S., Liu, Z. and Lu, Y. (2023) Temporal Data Meets LLM — Explainable Financial Time Series Forecasting.