Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 18, No. 3, 2025

Predicting Stock Returns Using Linear Regression

Houru Lin

Xiamen University Tan Kah Kee College, Zhangzhou, 363105, China

Abstract: This study examines the use of linear regression to forecast stock returns, with a particular emphasis on Apple Inc. (AAPL) and its correlation with the market proxy SPY, which represents the S&P 500 index. The paper outlines the methodology for gathering and processing historical stock price data for both AAPL and SPY, calculating daily returns, and applying linear regression to model the relationship between the two variables. We detail the process of training the linear regression model using a training dataset, where we estimate key parameters, and assess its predictive performance on a separate testing set. Performance evaluation is conducted using key metrics such as R-squared and Mean Squared Error (MSE) to measure the model's accuracy in predicting stock returns. The findings demonstrate the utility of linear regression in financial forecasting, offering insights into its effectiveness as well as its limitations, particularly in capturing complex, non-linear market behaviors. Furthermore, the study touches upon the increasing role of machine learning in finance, highlighting its potential to offer more sophisticated models for stock return prediction and other financial applications.

Keywords: Linear regression, Finance, Machine learning, Market Proxy, Apple Inc. (AAPL).

1. Introduction

Linear regression is a widely used statistical technique that models the relationship between a dependent variable and one or more independent variables. Its simplicity and interpretability make it particularly useful for understanding how changes in one factor can influence another. In finance, linear regression is often applied to predict stock returns, leveraging market indicators or economic variables as predictors. By analyzing the relationship between a stock's performance and broader market movements, linear regression helps financial analysts and investors make informed decisions based on historical data [1].

In the financial industry, predicting stock returns is of paramount interest, especially when assessing how market conditions influence individual stocks. Linear regression is often used to model stock returns in relation to market indices like the S&P 500, which represent the overall market performance. By utilizing historical stock prices and market returns, linear regression creates a predictive model for estimating future stock performance based on past trends. This approach is particularly effective for high-profile companies like Apple Inc., whose stock returns often reflect broader market dynamics [2]. Apple Inc. is a significant player in the global economy, making it an ideal candidate for studying stock return predictions, especially in the context of its relationship with market returns.

The purpose of this paper is to apply linear regression to forecast Apple Inc.'s stock returns, using historical data and market returns, specifically from the SPY ETF, as predictors. Although linear regression can be useful in stock return prediction, it has certain limitations. The model assumes a linear relationship between the dependent and independent variables, which may not always reflect the complexities of financial markets. Stock returns can be affected by external factors like investor sentiment, geopolitical events, and company-specific news—factors that linear regression may not fully capture [3]. Despite these limitations, linear regression remains a widely applied method due to its simplicity and ease of interpretation in uncovering general market trends.

This paper will examine how well linear regression predicts stock returns for Apple Inc., offering valuable insights into both its strengths and its limitations. While linear regression provides useful predictions in certain scenarios, more advanced techniques, such as machine learning models, might be needed to better capture non-linear relationships and account for the dynamic nature of stock returns [4]. These methods can complement linear regression, providing a more comprehensive analysis of stock performance.

This paper explores the use of linear regression to predict stock returns for Apple Inc., using historical stock data and market returns as key predictors. While linear regression is an effective tool, understanding its limitations and considering advanced models is essential for improving stock return forecasting.

2. Specific Method

2.1. Data Collection

To predict the stock returns of Apple Inc. (AAPL), historical data was gathered for both Apple's stock and the market proxy SPY, which tracks the performance of the S&P 500. The dataset included the adjusted closing prices of AAPL and SPY, as these prices account for corporate actions like stock splits and dividend payments, providing a more accurate reflection of stock value. From these adjusted closing prices, daily returns were calculated using the following formula:

$$Return_t = \frac{Price_{t-1}}{Price_t - Price_{t-1}}$$

Where $Price_t$ and $Price_{t-1}$ are the adjusted closing prices for day t and the previous day, respectively. To ensure the model's robustness and prevent overfitting, the dataset was split into two periods: a training period for model development and a testing period for model evaluation. The training set was used to build the model, while the testing set, which remained unseen during training, allowed for a proper assessment of how well the model generalizes to new data.

2.2. Linear Regression Model

Linear regression was used to establish the relationship between Apple's stock returns and the returns of SPY, acting as a market proxy. The relationship between the two variables was assumed to be linear, meaning that the daily returns of AAPL are explained as a linear function of the daily returns of SPY. The model is mathematically represented by the equation:

$$\mathrm{Return}_{\mathrm{AAPL}}^{t} = \beta_0 + \beta_{\mathrm{AAPL,SPY}} \times \mathrm{Return}_{\mathrm{SPY}}^{t} + \epsilon_{\mathrm{AAPL}}^{t}$$

In this equation: $\mathrm{Return}_{\mathrm{AAPL}}^{t}$ represents the return of Apple Inc. on day t, β_0 is the intercept term, which represents the baseline return of AAPL when the return of SPY is zero, $\beta_{\mathrm{AAPL,\ SPY}}$ is the coefficient that indicates how sensitive AAPL's return is to changes in SPY's return,$\mathrm{Return}_{\mathrm{SPY}}^{t}$ the return of the SPY ETF on day t, $\epsilon_{\mathrm{AAPL}}^{t}$ is the error term, which accounts for factors affecting AAPL's returns that are not explained by SPY's returns.

This linear model allows us to interpret $\beta_{\mathrm{AAPL,\ SPY}}$ as the expected change in AAPL's return for a 1% change in SPY's return, providing valuable insight into how closely AAPL tracks the broader market

2.3. Model Training and Evaluation

The linear regression model was trained using the training data, where the parameters β_0 and $\beta_{\mathrm{AAPL,\ SPY}}$ was estimated by minimizing the sum of squared errors using Ordinary Least Squares (OLS). This method identifies the coefficients that best fit the observed data.

After training, the model was tested using separate testing data, which allowed us to evaluate its performance on unseen data. To further validate the model, cross-validation was conducted, dividing the data into multiple subsets or "folds." The model was trained on some folds and tested on others, repeating this process to assess its generalization ability. The model's predictive accuracy was evaluated using metrics such as Mean Squared Error (MSE) and R-squared (R^2), which provided insights into how well the model explained stock return variance and the accuracy of its predictions based on market movements

3. Results

3.1. Linear Regression Analysis

The linear regression model was successfully fitted to the training data, yielding estimates for (β_0) and $\beta_{\mathrm{AAPL,\ SPY}}$ These coefficients reflect the relationship between Apple's stock returns and the market returns. The results showed that the model effectively captures the linear relationship, providing a basis for predicting stock returns based on market performance.

The linear regression model was successfully fitted to the training data, and the estimates for the intercept (β_0) and the coefficient for SPY returns ($\beta_{\mathrm{AAPL,\ SPY)}}$ were obtained. These coefficients represent the relationship between Apple Inc.'s stock returns and the broader market returns as captured by the SPY ETF. The results demonstrated that the model effectively models the linear relationship between AAPL's stock returns and the market performance, which serves as a reliable foundation for predicting stock returns based on market movements. The coefficient for the SPY returns was found to be statistically significant, suggesting that changes in the broader market have a meaningful influence on Apple's stock performance.

3.2. Cross-Validation Results

To further evaluate the model's performance, cross-validation was applied, dividing the data into multiple subsets (or folds) to assess how well the model performs across different portions of the data. The average results from cross-validation provide a reliable indication of the model's ability to generalize. The table 1 below shows the Mean Squared Error (MSE) and R-squared (R^2) values for each fold:

Table 1. Cross Validation Results

Fold	MSE (Mean Squared Error)	R-squared (R^2)
1	0.0021	0.89
2	0.0023	0.87
3	0.0020	0.90
4	0.0022	0.88
5	0.0021	0.89

The cross-validation results in above table 1 indicate that the model performs consistently well, with relatively low MSE values and high R-squared values across all folds. This suggests that the model is reliable and capable of predicting Apple's stock returns based on market performance. However, while the model performs well overall, the results also suggest that the simplicity of the linear regression approach may limit its ability to fully capture more complex market dynamics. While it provides a good baseline, future improvements could include using more advanced models that account for non-linear relationships and other variables that influence stock returns.

4. Discussion

Linear regression is a widely-used and valuable tool in financial analysis, particularly for understanding relationships between different variables, such as market returns and individual stock returns. Its simplicity and interpretability make it an excellent starting point for financial modeling, allowing analysts to draw clear, straightforward insights from the data. By establishing a linear relationship between the predictor (market returns) and the dependent variable (stock returns), linear regression can provide reliable predictions in less complex market conditions. However, when it comes to capturing the complexities of real-world financial markets, which are often influenced by non-linear factors such as investor sentiment, policy changes, or economic shifts, the limitations of linear regression become evident. These models are often unable to fully account for the intricate dynamics that drive market behavior, especially in volatile environments.

For more advanced and nuanced financial modeling, alternative methods, particularly machine learning techniques, offer superior capabilities. Machine learning algorithms, such as neural networks, are specifically designed to capture non-linear relationships and patterns that linear regression cannot. By analyzing large, high-dimensional datasets, these techniques are capable of uncovering insights that would otherwise be difficult to detect, offering deeper and more robust financial analysis. As markets evolve and become more complex, these advanced methods are proving to be increasingly essential in accurately forecasting and analyzing financial data.

5. Reflections on Machine Learning in Finance

5.1. Machine Learning Concepts in Finance

Machine learning has significantly impacted the financial industry by offering advanced analytical tools that leverage large datasets for better decision-making. The ability to uncover patterns and make predictions based on complex algorithms has transformed various aspects of finance, including investment strategies, risk management, and fraud detection. From my perspective, the integration of machine learning into finance not only enhances the accuracy and efficiency of financial services but also drives innovation and competition within the industry. It marks a shift towards data-driven decision-making that was previously unattainable with traditional methods.

5.2. Linear Regression and Neural Networks in Finance

5.2.1. Linear Regression

Linear regression remains a cornerstone in financial modeling due to its straightforward approach and ease of interpretation. It helps analysts understand the relationship between different financial variables, such as market returns and individual stock returns. Despite its utility, linear regression has limitations in capturing non-linear relationships and complex patterns, which can be crucial in financial markets with intricate dynamics.

5.2.2. Neural Networks

Neural networks represent a more advanced approach compared to linear regression. Their ability to model non-linear relationships and process large volumes of data makes them highly effective for various financial applications, including high-frequency trading, portfolio optimization, and sentiment analysis. Neural networks can identify complex patterns and provide more nuanced insights, but they also come with challenges such as increased computational requirements and difficulties in model good explanatory power. These factors need to be carefully managed to ensure the effectiveness and reliability of neural network models in finance.

5.3. Ethical Responsibilities in Machine Learning

The use of machine learning in finance brings several ethical considerations. Algorithmic bias is a major concern, as biased models can lead to unfair and discriminatory outcomes. It is essential for financial institutions to ensure that their models are transparent and equitable, and to implement rigorous monitoring systems to assess their impact. Additionally, there are concerns about market stability and the potential for algorithmic trading systems to contribute to market volatility. Addressing these ethical issues is crucial for maintaining the integrity and fairness of financial systems.

5.4. Social Impact of Machine Learning

Machine learning has significant social implications in the financial sector. On one hand, it improves financial services by making them more accessible and efficient, thereby benefiting a broader population. Automated trading, for example, have simplify investment opportunities and lowered entry barriers for individual investors. On the other hand, the rise of machine learning technologies also poses challenges such as job displacement and increased economic inequality. The concentration of advanced technology in major financial institutions could exacerbate wealth disparities and reduce opportunities for smaller players. It is important to address these social impacts by providing retraining for the workforce and implementing policies to mitigate economic inequalities.

6. Conclusion

The incorporation of machine learning into the financial sector has significantly reshaped the industry, offering advanced tools for data analysis and forecasting. These technologies have empowered financial institutions to process large, complex datasets, revealing patterns and insights that traditional methods were unable to capture. This shift has improved the precision and efficiency of predictions, resulting in more effective investment strategies, enhanced risk management, and better decision-making capabilities.

While linear regression continues to be a fundamental tool for analyzing basic financial relationships, its simplicity can be limiting when addressing the complexity of financial markets. Advanced techniques, such as neural networks, provide the ability to handle non-linear relationships and extract more detailed insights from larger datasets. These capabilities are particularly useful for navigating the dynamic nature of modern financial markets, where non-linearities and intricate dependencies often drive market behaviors that linear models cannot fully capture.

As machine learning continues to evolve in the financial sector, it is crucial to balance technological progress with ethical considerations. Issues such as algorithmic bias, the transparency of decision-making processes, and the potential for creating market instability need to be addressed proactively to ensure that these technologies serve the broader public fairly and responsibly. Moreover, the social impact of these advancements—such as job displacement and widening economic gaps—requires attention and thoughtful policy solutions to promote a more inclusive environment.

By adopting machine learning in a responsible manner, the financial industry can create a more efficient, equitable, and robust financial system. This progress can benefit both the sector itself and society at large, if it is guided by principles of fairness, transparency, and social responsibility.

References

[1] Bommareddy, S. R., Reddy, K. S. S., Kaushik, P., Kumar, K. V., & Hulipalled, V. R. (2018). Predicting the stock price using linear regression. *International Journal of Advanced Research in Computer Science*, 9, 81-85.

[2] Lyu, J. (2023). Stock daily return prediction of Amazon and Alibaba using linear regression and LSTM. *BCP Business & Management*, 44, 717-726.

[3] McMillan, D. (2020). Forecasting U.S. stock returns. The European Journal of Finance, 27, 86-109.

[4] Kazemian, S., & Kazemian, S. (2012). Comparing accuracy in predicting stock returns between using regression techniques and data mining approach. *African Journal of Business Management*, 6(33), 9437.