International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume:09, Issue: 11 "November 2024"

Enhancing Stock Return Predictions: Comparing Machine Learning Methods with Traditional Financial Models

Arav Agarwalla

The British School, New Delhi, India

DOI: 10.46609/IJSSER.2024.v09i11.016 URL: https://doi.org/10.46609/IJSSER.2024.v09i11.016

Received: 29 October 2024 / Accepted: 17 November 2024 / Published: 20 November 2024

ABSTRACT

This paper aims to understand if machine learning models can enhance stock price predictions compared to that of traditional financial models. The paper covers traditional financial models such as Stochastic Discount Factor models, factor-based models, option pricing models, and behavioural models, and machine learning techniques like supervised learning, deep learning, and hybrid models. By summarizing the results of various papers, this review compares the predictive accuracy of these models. The review found that machine learning methods, deep learning and hybrid models, outperformed traditional models as they captured nonlinear relationships between factors and stock price however found that some machine learning models were prone to overlearning. Through this review, financial analyst can understand if machine learning models should be used, which models to use specifically and can lead to enhanced stock price prediction.

Keywords: Machine Learning, Stock Price Prediction, Financial Models

1. Introduction

Machine Learning models are becoming increasingly prevalent in today's society and have a wide range of use-cases. They are able to learn and improve upon from past experiences and datasets, which makes them ideal for stock price predictions. Traditionally, for stock price prediction, linear models have been used, where data was regressed against factors that influence stock price. This review aimed to understand how can machine learning methods improve the prediction of stock returns compared to traditional financial models. By doing this review, we can develop a better understanding of how useful machine learning is compared to traditional models, when it comes to stock price prediction and can understand the specific models that should and shouldn't be used. One paper by Brian Kelly, *Empirical Asset Pricing via Machine Learning*, found that machine learning models provided more accurate predictions than traditional stock price prediction models (Kelly). Results showcasing that machine learning models were superior were also supported in the paper *Price Prediction of Share Market using Artificial Neural Network (ANN)* by Zabir Haider Khan. He found that ANN can effectively predict stock prices, with accuracy improving when a comprehensive set of financial indicators is used as inputs (Khan)

To conduct the review, first different types of traditional financial models will be looked at – stochastic discount factor and equilibrium models, factor models, behavioural and anomaly-based models, and then option based derivative models. Afterwards, different machine learning models will be looked at – supervised learning models, deep learning models, ensemble learning models and hybrid models. By looking at various papers we can get a general understanding of how each model performs and if there are some that work better than others and compare that to traditional financial models.

2. Traditional Financial Models

2.1. Stochastic Discount Factor (SDF) and General Equilibrium Models

SDF models are used to understand and forecast the pricing of stocks by evaluating how future cash flows should be discounted to their present value, while considering both time and risk. In asset pricing, the SDF reflects the market's risk preferences, ensuring that the price of an asset properly compensates investors for the level of risk associated with future cash flows.

In the 1997 paper *Assessing Specification Errors in Stochastic Discount Factor Models***,** Lars Peter Hansen and Ravi Jagannathan address the challenge of comparing asset pricing models when their implied stochastic discount factors (SDFs) do not perfectly price all assets, aiming to measure model misspecification (occurs when the assumptions or structure of a statistical or economic model do not accurately reflect the true underlying relationships). Their methodology introduces the Hansen-Jagannathan (HJ) distance, a metric that quantifies the distance between a model-implied SDF and the set of admissible SDFs that correctly price assets. They formulate two optimization problems: one that minimizes the variance of pricing errors without imposing positivity constraints on the SDF, and another that includes a positivity constraint to ensure no arbitrage opportunities. Applying this framework to various models—including the Capital Asset Pricing Model (CAPM) and consumption-based models—using time-series asset returns data, they find significant specification errors in commonly used models, especially linear factor models like the CAPM (Results). Their findings suggest that while no single model perfectly prices all assets, the Hansen-Jagannathan approach provides a nuanced comparison by highlighting pricing errors and emphasizing the importance of arbitrage-free conditions, thereby offering valuable insights into model performance (Hansen).

ISSN: 2455-8834

Volume:09, Issue: 11 "November 2024"

In the paper *Comparing Asset Pricing Models by the Conditional Hansen-Jagannathan Distance* by Patrick Gagliardini and Diego Ronchetti, they understand how to compare non-nested parametric specifications of the SDF using the conditional HJ distance. The research question asks how well different SDF models, including beta-pricing and preference-based models, match dynamic no-arbitrage restrictions for managed portfolios. Using a Generalized Method of Moments (GMM) estimator to compute the conditional HJ distance which accounts for a model's ability to match dynamic pricing restrictions across multiple conditions. They compare fourteen SDF models, such as the Fama-French models and the Capital Asset Pricing Model (CAPM), using this distance. The results demonstrate that the conditional HJ distance provides a more accurate comparison of model performance than traditional methods, particularly in highlighting differences in misspecification across models. This approach allows for a better understanding of how different factors impact pricing accuracy under dynamic conditions (Gagliardini and Ronchetti).

2.2. Factor Based Models

Factor-based models look at variables that are believed to influence not only stock returns but also broader asset prices and risks across different financial markets.

One critical paper in the study of factor models is William Sharpe's paper *Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk*, which addresses how capital asset prices are determined when investors face risk and explores the relationship between an asset's risk and expected return. Sharpe introduced the Capital Asset Pricing Model (CAPM) to explain how risk and return are related in market equilibrium.

Fig 1.0 – CAPM Model and Beta Formula

$$E(R_i) = R_f + \beta_i \left(E(R_m) - R_f\right)$$

Where:

- $E(R_i)$ = Expected return on asset i
- R_f = Risk-free rate of return
- β_i = The sensitivity of asset i to movements in the overall market (beta coefficient)
- $E(R_m)$ = Expected return of the market portfolio
- $E(R_m) - R_f$ = Market risk premium (the additional return expected from the market above the risk-free rate)

$$\beta_i = \frac{Cov(r_i, r_m)}{\sigma^2(r_m)}$$

ISSN: 2455-8834

Volume:09, Issue: 11 "November 2024"

In CAPM, the market factor represents the excess return of the market portfolio over the risk-free rate, and the model posits that an asset's expected return is determined by its sensitivity to this market factor. CAPM builds on earlier portfolio theory, assuming investors are rational, risk-averse, and aim to maximize returns while minimizing risk. It assumes investors choose between a risk-free asset and a market portfolio of risky assets. According to Sharpe, all investors share the same expectations about risk and returns, leading them to hold the market portfolio along with the risk-free asset. The CAPM formula shows that an asset's expected return is determined by its beta (β), which measures its sensitivity to market movements. Higher-beta assets, being riskier, require higher returns, while lower-beta assets offer lower returns. Sharpe's model suggests investors should be compensated only for systematic risk, as unsystematic risk does not influence returns (Sharpe)

While CAPM looked at one factor, in the paper, "*An Empirical Investigation of the Arbitrage Pricing Theory*" Richard Roll and Stephen A. Ross aimed to see if the Arbitrage Pricing Theory (APT) could provide a more valid explanation for the cross-section of expected stock returns than the single-factor Capital Asset Pricing Model (CAPM). APT is an asset pricing model that explains the relationship between an asset's expected return and multiple market factors, based on the principle that arbitrage will eliminate price discrepancies in efficient markets. Ross and Roll empirically tested APT, using daily data from New York and American Stock Exchange stocks from 1962 to 1972. Through factor analysis, they identified three or four significant risk factors that affect returns - changes in industrial production (real income), changes in interest rates, and changes in inflation rates. Cross-sectional regressions showed that expected returns were closely tied to factor loadings, with assets more sensitive to these factors yielding higher expected returns. Crucially, while systematic risks were the primary drivers of expected returns, idiosyncratic (firm-specific) risks had a minimal influence once systematic risks were considered, supporting APT's claim that only systematic risks affect asset pricing, contrasting with CAPM's single-factor model (Roll and Ross).

The idea that multiple factors affect returns was also studied by Fama and French. In their paper *Common Risk Factors in the Returns on Stocks and Bonds*, Fama and French extended the CAPM model by developing the Three-Factor Model. They aimed to determine whether common risk factors explain variations in returns for both stocks and bonds and whether these factors could form a unified model. The Three-Factor Model adds two factors to CAPM: a size factor (small vs. large firms) and a value factor (high vs. low book-to-market ratios). Using time-series regressions on 25 stock portfolios (sorted by size and book-to-market ratios), they tested the model's ability to explain returns. The results showed that the three-factor model significantly explains the variation in stock returns, with both the size and value factors being statistically significant. Smaller firms and those with higher book-to-market ratios tend to earn higher returns. However, bond returns were mainly influenced by term structure and default risk, making the model less effective for bonds than for stocks (Fama and French).

Carhart extended the Fama-French Three-Factor Model by adding momentum as a fourth factor in his study "*On Persistence in Mutual Funds*". He sought to determine whether past performance predicts future mutual fund returns and if performance persistence is due to systematic factors or manager skill. Momentum reflects the tendency of stocks that have performed well recently to continue doing so in the short term. Carhart tested the Four-Factor Model by sorting mutual funds into deciles based on past performance and regressing future returns against the market, size, value, and momentum factors. His findings revealed that the persistence of mutual fund performance was primarily driven by the momentum factor, rather than fund managers' skill. Funds with strong past performance typically held stocks with high momentum. However, once expenses and turnover costs were considered, few funds consistently outperformed the market on a risk-adjusted basis. Carhart concluded that past performance is not a reliable predictor of future returns because returns are largely driven by systematic risk factors and not persistent managerial skill (Carhart).

Later on, Fama and French, in their paper *A Five-Factor Asset Pricing Model*, extended their earlier Three-Factor Model by adding profitability and investment as new factors. They aimed to determine whether this five-factor model better explains stock returns than the original model. Using U.S. stock returns from 1963 to 2013, they tested the model through regression analysis, sorting portfolios by size, value, profitability (the return differential between stocks with high operating profitability and those with low operating profitability) and investment (the return difference between stocks of companies that invest aggressively in growth opportunities and those that invest more conservatively) to assess its explanatory power.. The results showed that the five-factor model outperformed the three-factor model, especially in explaining returns linked to profitability and investment. Firms with higher profitability and conservative investment policies tended to have higher returns, making both factors statistically significant. However, the value factor became redundant when the profitability and investment factors were included, indicating that these two factors captured much of the variation previously explained by value. Despite the model's improvement, it struggled to account for the returns of small, high-investment, low-profitability firms, leaving some gaps in explanation (Fama and French).

2.3. Option Pricing Models

Before this paper reviewed traditional financial models that looked at stock price prediction based on systematic risk factors. In this section the paper will shift the focus to derivatives, such as options, a more complex approach is required. Unlike stocks, which represent ownership in a company, options provide the right, but not the obligation, to buy or sell an asset at a predetermined price in the future.

Option pricing models are mathematical models used to determine the value of options, which are financial derivatives granting the right, but not the obligation, to buy or sell an underlying asset at a specified price before or at a certain date. These models consider various factors such as the current price of the underlying asset, the strike price, time to expiration, volatility, risk-free interest rates, and dividends to calculate the option's premium.

One of the most influential option pricing models is the Black-Scholes Model, developed by Black and Scholes in 1973. The Black-Scholes Model revolutionized financial economics by providing a theoretical estimate of the price of European-style options. It calculates an option's fair value based on factors such as the current price of the underlying asset, the strike price, time to expiration, risk-free interest rate, and volatility.

The research question of the 1982 paper "Option Pricing Model Estimates: Some Empirical Results" by Gultekin, Rogalski, and Tinic was to assess how accurately the Black-Scholes model estimates option prices and to explore the factors affecting pricing biases. They used data from call options traded on the Chicago Board Options Exchange (CBOE) from 1975 to 1976, applying the Black-Scholes formula to estimate premiums over several weeks before expiration. The actual option prices were compared with the estimated premiums, and biases were tested against variables like stock volatility, time to expiration, and whether options are "in-the-money" or "out-of-the-money" The study found that the Black-Scholes model underestimates premiums for high-volatility stocks and "out-of-the-money" options, and overestimates prices for low-volatility stocks and "in-the-money" options. Additionally, the errors decrease as the time to expiration shortens. These findings indicate that the Black-Scholes model has systematic biases related to volatility, moneyness, and time to expiration, highlighting limitations in the model's assumptions under certain market conditions.

2.4. Behavioural and Anomaly-Based Models

Behavioural models in finance seek to explain how psychological factors and cognitive biases influence the decisions of investors and the behaviour of markets. Unlike traditional finance theories, which often assume rational behaviour, behavioural finance recognizes that investors are not always rational, and that emotions, heuristics, and social factors can lead to errors in judgment. Anomaly-based models focus on empirical phenomena that deviate from expected outcomes predicted by classical financial theories.

In the paper, *"Investor Sentiment and the Cross-Section of Stock Returns"*, Malcolm Baker and Jeffrey Wurgler aimed to explore how investor sentiment influences the cross-section of stock

ISSN: 2455-8834

Volume:09, Issue: 11 "November 2024"

returns, focusing on firms that are difficult to value and arbitrage. They hypothesized that stocks such as small, young, high-volatility, unprofitable, non-dividend-paying, and distressed firms would be more sensitive to investor sentiment. Using monthly stock return data from 1963 to 2001, the authors constructed a composite sentiment index based on six proxies, including the closed-end fund discount and IPO activity. They examined how stock returns across decile portfolios sorted by firm characteristics varied depending on the level of sentiment at the beginning of each period. The results showed that when sentiment was low, these stocks earned higher returns, while during high-sentiment periods, they performed worse. This suggests that investor sentiment leads to mispricing, especially for stocks with subjective valuations, challenging the traditional view that only systematic risks determine returns (Baker and Wurgler)

2.5. Evaluation

Traditional financial models for stock prediction do have their advantages. They provide structured frameworks for assessing risk and expected returns, allowing investors to make informed decisions. They are able to map returns to some extent, which can be useful when assessing performance and predicting stock returns. However, their predictive accuracy can be limited as they often rely on oversimplified assumptions, such as constant risk premiums and linear relationships. This can lead to inaccurate predictions as markets are volatile and dynamic. Moreover, the static nature of traditional models may hinder their effectiveness in adapting to changing economic conditions or incorporating new data types.

3. Machine Learning Models

Machine learning models can be categorized into 2 categories – Supervised Learning and Unsupervised Learning. Supervised learning involves training a model on a labelled dataset, where both the input data and the corresponding correct outputs (labels) are known. With Unsupervised Learning, the model works with data that has no labelled output, meaning that it must learn the patterns or structure directly from the input data.

3.1. Supervised Learning Methods

Supervised learning methods refer to machine learning techniques where models are trained on labelled datasets. Neural Networks are a type of supervised learning models that are computational models that mimic the complex functions of the human brain. The neural networks consist of interconnected nodes or neurons that process and learn from data, enabling tasks such as pattern recognition and decision making in machine learning. Support Vector Machines and data mining techniques are also types of supervised learning, which enable accurate predictions by identifying patterns and relationships in data. Support Vector Machines

ISSN: 2455-8834

Volume:09, Issue: 11 "November 2024"

are effective for classification and regression tasks by finding decision boundaries, while data mining techniques involve finding insights from large datasets to inform decision-making.

One paper that looks at this model is *Stock Price Prediction Using Support Vector Machine Approach* by Naliniprava Tripathy. His research question was "Can Support Vector Machine (SVM) accurately predict the direction of stock price movements, specifically for the S&P BSE TECK index, using historical data from 2008 to 2018?" The study used the SVM model to predict the daily directional movement of the S&P BSE TECK index over a 10-year period. They inputted the stock price volatility, stock momentum, index volatility, and index momentum, into the model. The model classified the price movements as either up or down (binary movements), making it a classification problem (a type of task where the goal is to assign inputs into predefined categories). The Radial Basis Function (RBF) kernel (a mathematical function used in Support Vector Machines to transform data into a higher-dimensional space) is used in the SVM to handle non-linear boundaries (complex decision surfaces that cannot be separated with a straight line). The performance of the model is measured using the Hit Ratio, which calculates the percentage of correct predictions of stock price direction compared to actual results. The study found that the SVM model achieved an average prediction accuracy of 60.2% after the 2008 financial crisis. The model's accuracy was highest for short-term predictions, with lower accuracy for long-term forecasts. The study concluded that the SVM model can be used effectively for short and medium-term stock price prediction in the Indian stock market (Tripathy).

Another paper that examines supervised learning models for stock price prediction is "Predicting Stock Prices Using Data Mining Techniques" by Qasem A. Al-Radaideh, Adel Abu Assaf, and Eman Alnagi. Their research question was, "Can data mining techniques, specifically decision trees, predict stock price movements based on historical data?" The study analysed stock data from three companies listed on the Amman Stock Exchange (ASE) from 2005 to 2007, using decision trees to classify whether an investor should buy or sell. They used the C4.5 and ID3 algorithms to construct decision trees, followed by pruning techniques to enhance the models. The performance was evaluated using K-Fold Cross Validation and the Holdout Method. The results showed classification accuracy ranging from 44% to 52%, depending on the company and the algorithm. While the decision trees provided useful predictions, external market conditions, such as political events, affected accuracy (Alnagi).

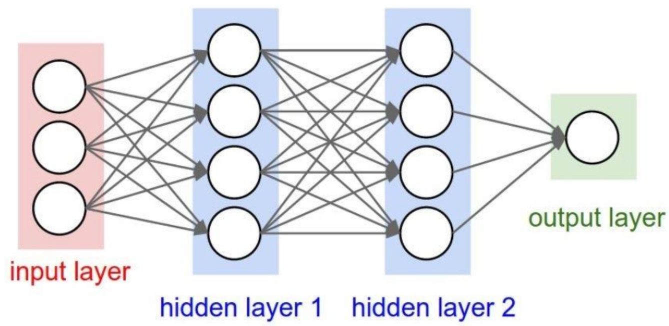
3.2. Deep Learning Models

A deep learning model is a type of machine learning model that utilizes artificial neural networks with multiple layers

ISSN: 2455-8834

Volume:09, Issue: 11 "November 2024"

Fig 2 - Image of a Deep Learning Model Architecture (Image taken from Towards Data Science)



A paper that looks at deep learning models is *"NSE Stock Market Prediction Using Deep-Learning Models*" by Hiransha Ma, Gopalakrishnan E.Ab , Vijay Krishna Menonab, Soman K.P. They looked at the ability of deep learning models to predict stock prices. Their research question was, "Can deep learning models predict stock prices from both NSE and NYSE stock exchanges using historical data?" and looked at four deep learning architectures: Multilayer Perceptron (MLP), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN). Using day-wise closing prices of companies from the NSE and NYSE, the models are trained on Tata Motors' stock data and tested on other companies. The models were compared to the ARIMA model. They found that CNN outperformed other models in capturing stock price patterns. They also found that models trained on NSE data could also successfully predict NYSE stock prices, suggesting that there are common factors between the markets. Results suggest that deep learning models outperform linear models like ARIMA in stock market prediction (Ma).

Building on the deep learning prediction frameworks for stock prices, another comprehensive study is the *"Financial Time Series Forecasting with Deep Learning: A Systematic Literature Review (2005-2019)"* by Sezer, Gudelek, and Ozbayoglu. Their research question was, "How have deep learning models evolved for financial time series forecasting, and how do they compare with traditional machine learning approaches?" The paper reviews over 150 studies from 2005 to 2019, focusing on the application of various deep learning models—like LSTM, CNN, RNN, and hybrid approaches—for forecasting financial markets, including stocks, forex, commodities, and cryptocurrencies. The study categorized models by their implementation areas and evaluation metrics (e.g., RMSE, MSE). Results revealed that LSTM and CNN-based models consistently outperformed traditional methods across different financial datasets, with LSTM excelling in temporal data tasks. The review highlighted both the progress and challenges in applying deep learning to finance, particularly in model complexity and feature selection (Sezer).

Another paper that looks at this is *Price Prediction of Share Market using Artificial Neural Network (ANN)* by Khan, Alin, and Hussain, the research question explored whether ANN can provide more accurate stock price predictions than traditional methods. The authors used a Feedforward Neural Network trained via backpropagation with five input variables: General Index, Net Asset Value (NAV), Price-to-Earnings (P/E) Ratio, Earnings Per Share (EPS), and Share Volume. The data was normalized for compatibility, and historical data from the Bangladesh Stock Exchange was used for training. Two simulations were conducted: one with 2 inputs and another with 5 inputs to assess the impact of input variables on prediction accuracy. The results showed that using more input variables significantly reduced prediction error. In the first simulation with 2 inputs, the error was 3.71%, while in the second simulation with 5 inputs, the error dropped to 1.53%. The sum of squared errors (SSE) during training further demonstrated the benefits of incorporating more variables, though the improvement diminished with more than 4 inputs. The study concluded that ANN can effectively predict stock prices, with higher accuracy achieved when more financial indicators are used (Khan).

3.3. Ensemble Learning Methods

Ensemble learning involves combining multiple models, of the same type to improve predictive performance.

One paper that looks at ensemble models is *"Predicting the Direction of Stock Market Prices Using Random Forest"* by Luckyson Khaidem, Snehanshu Saha, and Sudeepa Roy Dey, which addresses the question, "Can random forest classifiers predict stock price movement more effectively than traditional methods?" The study investigates historical stock market data using various technical indicators like the Relative Strength Index (RSI), which measures the speed and change of price movements to identify overbought or oversold conditions, and stochastic oscillators, which compare a particular closing price to a range of prices over time to predict potential reversals. The researchers pre-processed the data with exponential smoothing, a technique that reduces noise by giving more weight to recent observations. They then trained a random forest model. Their results, validated with Out-of-Bag (OOB) error estimates, a method that estimates error using observations not included in each tree's bootstrap sample, showed a significant reduction in error rates as more trees were added. The random forest classifier achieved accuracy rates between 84% and 94% for predictions ranging from 1 to 3 months, outperforming traditional methods like SVM and logistic regression. This study demonstrated the effectiveness of ensemble learning methods for stock prediction and highlighted random forest's superiority in long-term forecasting (Khaidem).

Building on previous efforts comparing machine learning methods for stock prediction, *"A Comprehensive Evaluation of Ensemble Learning for Stock Market Prediction"* by Isaac Kofi

ISSN: 2455-8834

Volume:09, Issue: 11 "November 2024"

Nti, Adebayo Felix Adekoya, and Benjamin Asubam Weyori offers a detailed examination of ensemble techniques like bagging, boosting, stacking, and blending. Their research question was "Which ensemble techniques are most effective across different global stock indices for both classification and regression tasks". Using data from the Ghana, Johannesburg, New York, and Bombay stock exchanges (2012–2018), 25 ensemble models were developed, leveraging Decision Trees (DT), Support Vector Machines (SVM), and Neural Networks (NN). Stacking consistently outperformed other methods, achieving 100% accuracy for both classification and regression, especially with DT and NN as base learners. Blending also performed well, but its computational cost was higher. Bagging and boosting, while effective, showed greater variance and lower performance compared to stacking, especially on smaller datasets like JSE. The study concluded that the choice of ensemble technique, dataset origin, and number of base learners significantly impact prediction accuracy (Nti).

3.4. Hybrid Methods

Hybrid machine learning models are models that make use of multiple different types of machine learning techniques, to leverage their strengths.

One paper that looks at a hybrid model is *"Framework for Predicting and Modeling Stock Market Prices Based on Deep Learning Algorithms"*, by Theyazn H. H. Aldhyani and Ali Alzahrani. Their research question was, "Can LSTM and CNN-LSTM models effectively predict stock prices for companies like Tesla and Apple using past data?" LSTM models are a type of recurrent neural network designed to model sequential data and capture long-term dependencies by using memory cells to retain information over time. CNN-LSTM is a hybrid model that combines convolutional neural networks for feature extraction from spatial data with LSTM networks for learning temporal patterns in sequential data. The study uses daily stock prices of Tesla (2014–2017) and Apple (2010–2020) to train the models. They applied both LSTM and a hybrid CNN-LSTM model to predict future prices. The performance was measured using metrics like MSE, RMSE, and R^2. Results showed that CNN-LSTM outperformed LSTM with R^2 values of 99.58% for Tesla and 99.87% for Apple during training. In testing, CNN-LSTM achieved better accuracy with R^2 values of 99.26% for Tesla and 99.73% for Apple. The hybrid CNN-LSTM was superior in forecasting and error reduction compared to the LSTM-only model (Alzahrani).

One paper that looks at asset pricing through machine learning is – *Empirical Asset Pricing via Machine Learning''* by Shihao Gu, Bryan Kelly, and Dacheng Xiu. Their research question was can machine learning techniques improve the accuracy of asset return predictions compared to traditional models in empirical asset pricing? The authors compared several machine learning models, including linear regressions, penalized regressions and nonlinear models like decision

ISSN: 2455-8834

Volume:09, Issue: 11 "November 2024"

trees, random forests, and neural networks, against traditional linear models like OLS and the Fama-French models. They use a dataset of nearly 30,000 stocks from 1957 to 2016, with over 900 predictive variables covering stock characteristics, time-series data, and industry dummies. The models are trained and evaluated using out-of-sample R^2 to assess predictive accuracy. They found that machine learning models, particularly nonlinear methods like decision trees and neural networks, significantly outperform traditional models in predicting stock returns. The out-of-sample R^2 for machine learning models ranges from 0.16% to 0.40%, compared to lower scores for OLS and Fama-French models. Additionally, neural network-based portfolio strategies demonstrate higher Sharpe ratios, indicating better risk-adjusted returns. The results show that machine learning can capture complex, nonlinear relationships in financial data, offering significant improvements in predictive accuracy and economic gains over traditional asset pricing models (Kelly).

3.5. Evaluation

The use of machine learning models in stock pricing has allowed for advanced techniques that enhance predictive accuracy and decision-making processes. Unlike traditional models, which often rely on linear assumptions and historical data, machine learning algorithms can analyse larger datasets and identify complex, non-linear relationships that may influence stock prices. Techniques such as supervised learning, deep learning, and ensemble methods allow for meaningful patterns to be showcased. Additionally, machine learning models can adapt to changing market conditions, allowing for real-time predictions that can improve trading strategies. However, the effectiveness of these models is contingent on the quality of data. In addition, there is the risk of overfitting and over learning when overly complex models are employed. Plus, these models still may not fully account for other factors, like news and market certainty factors.

4. Conclusion

The aim of this review was to see how can machine learning methods improve the prediction of stock returns compared to traditional financial models. This review found that traditional factor models were useful when it came to the prediction of stock returns and found that more factors added in these models did impact the explanatory power of the model. However, it was seen that these models failed to capture complex data structures. In contrast, machine learning approaches showcased superior predictive accuracy by using vast datasets and were able to identify nonlinear relationships. One shortcoming that was observed was that some models did have the tendency to overlearn hindering the accuracy. The findings suggest that integrating machine learning techniques could enhance stock price predictions. By using a variety of models, this review can showcase how different types of machine learning and traditional models compare.

ISSN: 2455-8834

Volume:09, Issue: 11 "November 2024"

However, this review may not be entirely comparable due to different datasets, like data from the New York Stock Exchange while others its certain stocks like Tesla and Apple, being applied, and different markets having different impacting factors. Future studies can employ a larger variety of studies and models to draw more conclusions, and make use of analytical techniques, to draw their own techniques. In conclusion, employing more advanced machine learning techniques in stock return prediction not only enhances accuracy but also positions researchers to better understand the complexities of financial markets by understanding the non-linearities between factors and stock prices, relationships that often were underexplained by traditional models.

References

- Hansen, Lars Peter, and Ravi Jagannathan. "Assessing Specification Errors in Stochastic Discount Factor Models." *The Journal of Finance*, vol. 52, no. 2, 1997, pp. 557–590.
- Gagliardini, Patrick, and Diego Ronchetti. *Comparing Asset Pricing Models by the Conditional Hansen-Jagannathan Distance*. May 2015, SSRN
- Sharpe, William F. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *The Journal of Finance*, vol. 19, no. 3, 1964, pp. 425–442.
- Roll, Richard, and Stephen A. Ross. "An Empirical Investigation of the Arbitrage Pricing Theory." *The Journal of Finance*, vol. 35, no. 5, 1980, pp. 1073–1103.
- Fama, Eugene F., and Kenneth R. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, vol. 33, no. 1, 1993, pp. 3–56.
- Carhart, Mark M. "On Persistence in Mutual Fund Performance." *The Journal of Finance*, vol. 52, no. 1, 1997, pp. 57–82.
- Fama, Eugene F., and Kenneth R. French. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics*, vol. 116, no. 1, 2015, pp. 1–22.
- Gultekin, N. Bulent, Richard J. Rogalski, and Sinan H. Tinic. "Option Pricing Model Estimates: Some Empirical Results." *The Journal of Finance*, vol. 37, no. 3, 1982, pp. 585–602.
- Baker, Malcolm, and Jeffrey Wurgler. "Investor Sentiment and the Cross-Section of Stock Returns." *The Journal of Finance*, vol. 61, no. 4, 2006, pp. 1645–1680.

ISSN: 2455-8834

Volume:09, Issue: 11 "November 2024"

- Tripathy, Naliniprava. "Stock Price Prediction Using Support Vector Machine Approach." *International Journal of Engineering and Advanced Technology*, vol. 8, no. 6, 2019, pp. 4858–4864.
- Al-Radaideh, Qasem A., Adel Abu Assaf, and Eman Alnagi. "Predicting Stock Prices Using Data Mining Techniques." *International Arab Journal of Information Technology*, vol. 8, no. 2, 2011, pp. 164–170.
- Ma, Hiransha, Gopalakrishnan E. Ab, Vijay Krishna Menonab, and Soman K. P. "NSE Stock Market Prediction Using Deep-Learning Models." *Procedia Computer Science*, vol. 132, 2018, pp. 1351–1362.
- Sezer, Omer Berat, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. "Financial Time Series Forecasting with Deep Learning: A Systematic Literature Review: 2005–2019." *Applied Soft Computing*, vol. 90, 2020, 106181.
- Khan, Zabir Haider, Tasnim Sharmin Alin, and Md. Akter Hussain. "Price Prediction of Share Market Using Artificial Neural Network (ANN)." *International Journal of Computer Applications*, vol. 22, no. 2, 2015, pp. 1–5.
- Khaidem, Luckyson, Snehanshu Saha, and Sudeepa Roy Dey. "Predicting the Direction of Stock Market Prices Using Random Forest." *Procedia Computer Science*, vol. 89, 2016, pp. 516–523.
- Nti, Isaac Kofi, Adebayo Felix Adekoya, and Benjamin Asubam Weyori. "A Comprehensive Evaluation of Ensemble Learning for Stock Market Prediction." *Journal of Information and Knowledge Management Systems*, vol. 50, no. 2, 2020, pp. 163–181.
- Aldhyani, Theyazn H. H., and Ali Alzahrani. "Framework for Predicting and Modeling Stock Market Prices Based on Deep Learning Algorithms." *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 4, 2021, pp. 4639–4655.
- Gu, Shihao, Bryan Kelly, and Dacheng Xiu. "Empirical Asset Pricing via Machine Learning." *The Review of Financial Studies*, vol. 33, no. 5, 2020, pp. 2223–2273.

