A Comprehensive Review of AI-Driven Techniques for Predicting Stock Market Trends

Geetika Chatley[1*], Apurva Tripathi[2], Aman Hayat[3], Aayush Mandilwar[4], Tallapelli Sharanaya[5] and Muskan Kumari[6]
{geetika.25107@lpu.co.in[1], apurvatripathi2105@gmail.com[2], amanhayate@gmail.com[3], aayushh158@gmail.com[4], akhileshthallapelli@gmail.com[5], muskankumari381@gmail.com[6]}

Assistant Professor, Department of Computer Science and Engineering, Lovely Professional University, Jalandhar, Punjab, India.[1]
Department of Computer Science and Engineering, Lovely Professional University, Jalandhar, Punjab, India [2, 3, 4, 5,6]

Abstract. Stock market prediction is essential in providing investors and companies with data-driven financial decisions. An accurate prediction of stock prices significantly enhances profit potential while reducing investment risks, making it invaluable in modern finance. This paper explores the application of Linear Regression, widely adopted and fundamental machine learning algorithm, for predicting stock prices. Linear Regression assumes that there is a linear relationship of the stock price (response variable) with several variables that could be independent like trading volume, historical prices, and other market-based indicators. The algorithmic approach seeks to fit a line of best fit which minimizes the discrepancy between the calculated and actual stock prices giving a model that is as simple as it is computationally efficient and interpret-able. Linear Regression provides easy clearances on how each factor contributes towards price movement. However, in complex patterns or with non-linear trends that might prevail in financial data sets, the performance of a model may be limited due to these features. Authors in this paper will have aimed at enhancing the prediction of the model by including several feature engineering techniques applied on the input variables followed by their refinement. More so, we outline how Linear Regression may be combined with other more advanced techniques, such as moving averages or sentiment analysis, to fine-tune the outcome even further. The paper goes on to enumerate the merits and demerits of Linear Regression, and then provides helpful guidance to investors for its usage in stock market prediction.

Keywords: Stock Market Prediction, Financial Forecasting, Investment Analysis, Stock Price Prediction, Machine Learning in Finance, Time Series Analysis

1 Introduction

Basic to financial analysis, stock market prediction is very crucial in guiding investments. It is the ability to predict stock price and fluctuation in the future, based on past data and various influencing factors. This is very important information that investors, financial analysts, and traders seek to get more informed decisions. Accurate stock market prediction leads to financial gains while wrong predictions lead to losses. The problem is that financial markets are dynamic and unpredictable with regard to the impact of innumerable variables, macroeconomic indicators, geopolitical events, investor sentiment, and firm-specific news.

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2358119

Traditionally, stock market forecasting mainly used technical analysis, as investors analyzed price charts and patterns to predict future directions. Although there have been changes in these areas of research, one revolution was brought by ML. Such models have begun to analyse vast volumes of data and look for patterns and trends that might have not been possible for the naked eye. The simple and more complex models such as linear, neural networks and so many others are continually applied to the prediction in the stock market since earlier methods would learn from what has occurred before and the ability to generalize them as predictions.

Among the different techniques of machine learning, we have regression, and perhaps the one that stands tall among other algorithms is due to being easy to design, with an intuitive interface, where interpret-ability gives it preference. This approach is named linear regression, used in predicting a dependent variable-- here, the value of stock--as correlated with historical prices, quantities traded in the market or other general economic indicators. Its simplicity enables an easy understanding of the relationships among the variables and has been preferred by beginners and practitioners alike for financial forecasting on account of its simplicity. The algorithm fits a linear equation to the observed data and attempts to get as small a difference between the predicted and actual values as possible. Despite its benefits, the application of linear regression in the prediction of stock prices suffers from a significant drawback. One of the reasons is that it inherently assumes a linear relationship between variables, which is not true in real-world financial markets all the time. Stock prices have a wide range of intricate, nonlinear factors like the sentiment of investors, macroeconomic fluctuations, and other unexpected changes in the market on which linear regression models have difficulty representing. Further, the stock market is also highly volatile, and so the data might contain noises, which causes bad models of prediction or over-fitting where a model would have performed exceptionally well on a given history but fail when exposed to unseen data. Moreover, a huge number of variables determine financial markets, many of which may interact with each other in non-linear ways. That makes it difficult to capture the full complexity of the market using linear regression models. For example, linear regression perfectly models simple relations between a limited number of variables but faces difficulties while trying to comprehend the subtle interactions between so many economic indicators, technical and investor factors. Furthermore, stock prices have temporal correlations, whereby past prices result in future prices that are simply not very easy to reproduce using straightforward linear models. The most recent trends are due to the increasing adoption of more advanced machine learning algorithms, such as decision trees, support vector machines, and deep learning models. For example, deep learning algorithms can better deal with issues of complexity and non-linearity in the financial marketplace. However, even amidst all these, linear regression will remain popular due to greater transparency, efficiency, computational ease, and easy explanation. It forms a basis for many financial forecasting studies because it contains useful information in terms of the relations even though lacking the sophistication of more complex models.

This paper is intended to be a comprehensive review of the application of linear regression in stock market prediction, outlining its strengths, challenges, and comparison with other machine learning methods. This review will elucidate the utility of this model in financial forecasting and its place in the broader landscape of machine learning approaches to stock market prediction by exploring the performance of linear regression under different market conditions and datasets.

2 Review of Literature

Stocks' prediction at the stock market has gained a lot of attention and has been highly published using different machine learning techniques and deep learning approaches. Patel et al. (2015) have proposed an SVMs and neural networks comparison including trend deterministic data preparation. LSTM networks were shown to be effective by Fischer and Krauss (2018), and deep learning models' pros and cons were discussed by Yu and Yan (2020). Chong et al. (2017) discussed deep learning's better performance capturing nonlinearities in financial data.

Support vector machines have been demonstrated to have better predictive ability than linear models (Kim, 2003; Huang et al., 2005; Chen & Chen, 2008). Ensemble techniques such as random forests improve accuracy (Kumar & Thenmozhi, 2016; Shah et al., 2019). Niu et al. (2020) Hybrid CNN-RNN was adopted by Ju et al.

deep learning, which also set a precedent for stock trend prediction. Feature selection studies for sentiment analysis have led to improvements in their model performance (Tsai & Hsiao, 2010; Wang & Vergara, 2016).

Relatively, ARIMA has been compared with deep learning's forecasting models and demonstrated that AI-based approaches do well than the traditional forecasting (Adebiyi et al., 2014; Zhong & Enke, 2017). Some hybrid models combining the prediction of regression and machine learning models have also been studied (Hsu, 2011; Huang & Peng, 2020). Feelings-enhanced models and technical analysis techniques have been proposed to enhance stock price forecasting (Wang & Vergara, 2016; Chong & Ng, 2008).

Other works like Ballings et al. (2015) and Atsalakis & Valavanis (2009) have studied the behavior of classifiers, and Tay & Cao (2001) and Chen et al. (2008) used support vector machines in finance. And the development of advanced hybrid models (Huang and Peng, 2020; Hsu, 2011) and deep learning-based models (Zhang et al., 2016; Niu et al., 2020) has further improved the accuracy of forecasting. Re-searh 1-indicators that may add meaning to MACD and RSI and may provide lead/lag relationship signals for technical analysis (Chong Ng, 2008).

In general, classical methods, e.g., linear regression and ARIMA, are the based methods, but the M L and deep learning including SVMs, RF, and NNs could have better predicting accuracy and more adaption to the complicated financial markets.

3 Methodologies in AI-Driven Stock Market Prediction

3.1 Machine Learning Models

Classical machine learning models such as SVM, Random Forest and Gradient Boosting Machine are commonly applied to stock market prediction. These approaches exploit the previous information to predict the future but they rely heavily on feature engineering. Random Forests tend to work well when faced with complicated data sets with a lot of variables but might not optimally capture time patterns within stock data.

3.2 Deep Learning Models

Deep learning models like RNNs and LSTMs are particularly useful in time-series prediction problems. Long short-term memory (LSTM) networks that can maintain long-term influences in the network topology become a good candidate for sequential feature extraction for stock predictions. Latently, each of the recent years has seen the boom of Convolutional Neural Networks (CNNs) [13]– [19]. also been applied to learn local patterns over stock data, and have shown promising results.

3.3 Natural Language Processing (NLP)

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{y_i})^2} \tag{1}$$

NLP techniques are increasingly used to analyze financial sentiment from news articles, social media, and financial reports. Sentiment Analysis, particularly,

$$R-squared\ (R^2)=\frac{\sum_{i=1}^{n}(y_i-\widehat{y_i})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{2}$$

helps gauge market sentiment, providing insight into potential stock movements. The BERT model, among other NLP models, is instrumental in understanding and quantifying the sentiment of unstructured data sources like tweets and news headlines.

- Indicates the proportion of variance in the target variable explained by the model.

Reinforcement learning, in which an agent learns optimal trading actions through trial and error, is gaining traction for its adaptability. Algorithms like Deep Q- Networks (DQN) and Policy Gradient Methods can optimize trading strategies based on real-time feedback, making them effective in environments with changing dynamics.

3.4 Performance Evaluation:

Assess the accuracy and robustness of the model using statistical metrics like Root Mean Square Error (RMSE) and R-squared, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE).

Here's how we can calculate all these values. Metrics for Performance Evaluation Mean Absolute Error (MAE)

Measures the average absolute difference between predicted and actual stock prices. Formula:

$$\boldsymbol{MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|} \qquad (3)$$

Mean Squared Error (MSE)

Calculates the average squared difference between predicted and actual values, penalizing larger errors.

Formula:

$$\boldsymbol{MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^{2}} \qquad (4)$$

Root Mean Squared Error (RMSE)

• The square root of MSE, providing an error metric in the same unit as the target variable.
Formula:

$$\boldsymbol{RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}} \qquad (5)$$

R-squared (R^2)

• Indicates the proportion of variance in the target variable explained by the model.
Formula:

$$\mathrm{R}^2 = \mathbf{1} - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (6)$$

4 Applications of AI in Stock Market Prediction

4.1 Stock Price Prediction

What can an AI model predict? These predictions are helpful to the investors in order to make out when to buy/sell the stocks. Though short-term prediction can be addressed by traditional methods such as LSTM, long-term prediction usually combines several models to decrease the error.

4.2 Portfolio Optimization

AI assists in portfolio balancing by forecasting risks and returns of individual stocks. Algorithms such as Markowitz Optimization, are today factorized with machine learning, using not only historical data, but also real-time data to optimize portfolios.

4.3 Risk Management

Leading edge AI models help in acknowledging and controlling financial risks as they are constantly analyzing the portfolio exposure and users can pre-emptively act upon imminent downturns.

4.4 Algorithmic Trading

AI makes it possible to trade automatically based on trading strategies that react to market changes in fractions of a second. Probably one of the most visible applications of AI is High Frequency Trading (HFT), which is based on advanced algorithms and leverages the ability to place huge amounts fraction of a second. Fig 1 shows MRF Ltd. All-time stock price chart.

Root Mean Squared Error (RMSE)

• The square root of MSE, providing an error metric in the same unit as the target variable.

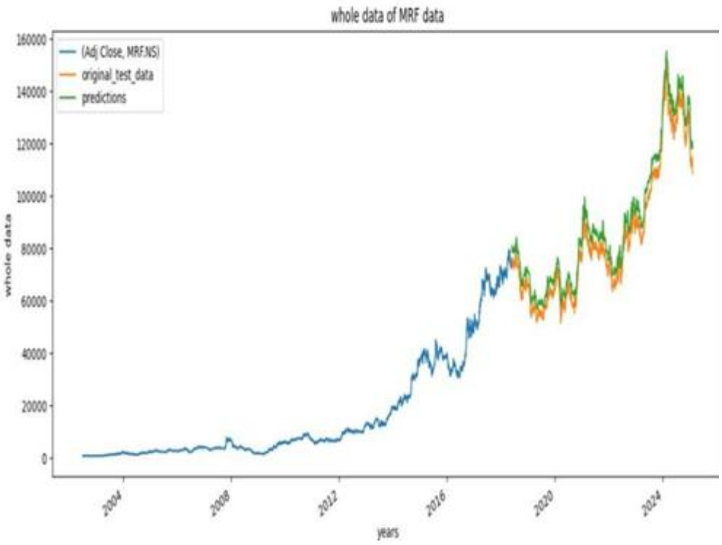

Fig. 1. MRF Ltd. All-time stock price chart.

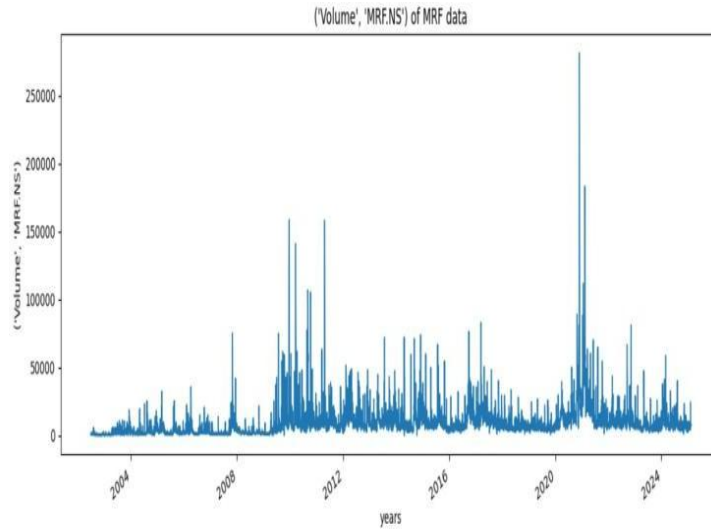

Fig. 2. Volume of MRF All-time data.

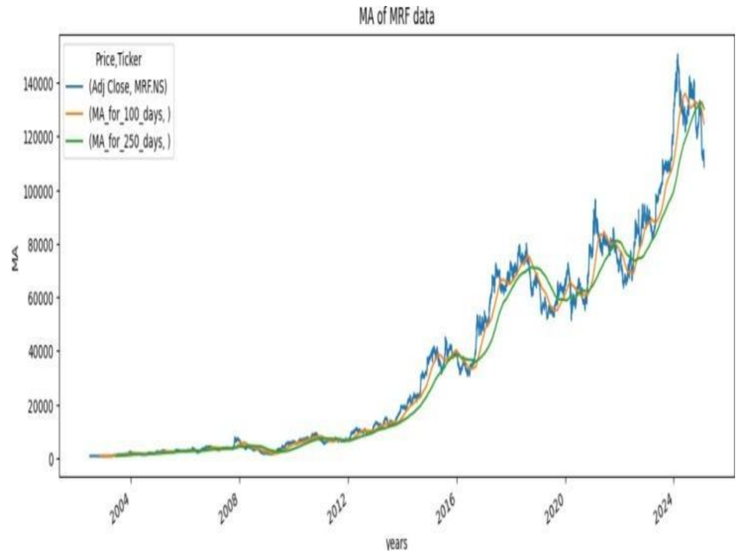

Fig. 3. Moving-Average of MRF All-time stock data.

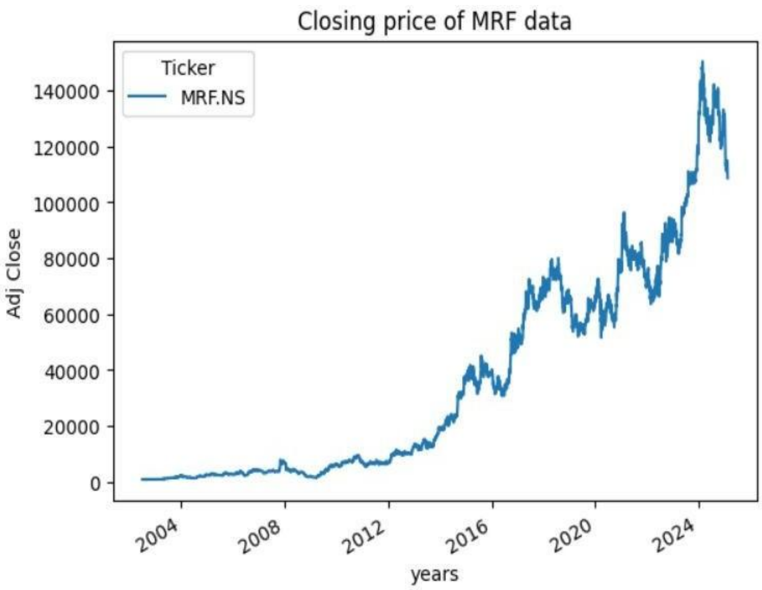

Fig. 4. Closing price of MRF Data.

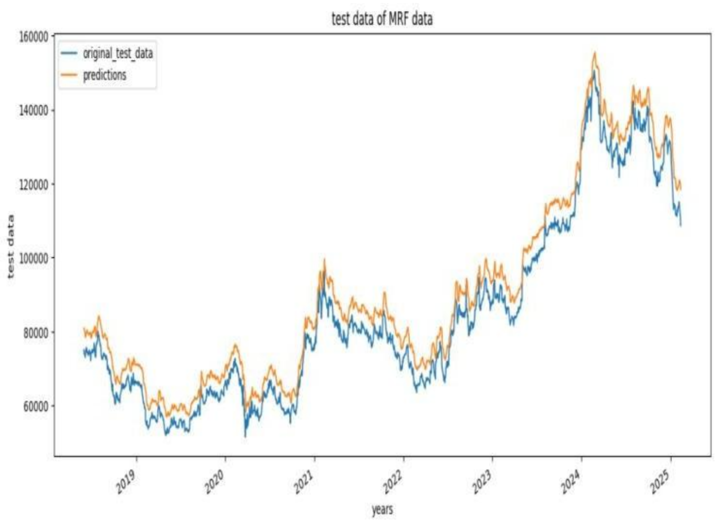

Fig. 5. Test data of MRF All-time Stock Price data.

```
model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 100, 128)	66,560
lstm_1 (LSTM)	(None, 64)	49,408
dense (Dense)	(None, 25)	1,625
dense_1 (Dense)	(None, 1)	26

Total params: 352,859 (1.35 MB)

Trainable params: 117,619 (459.45 KB)

Non-trainable params: 0 (0.00 B)

Optimizer params: 235,240 (918.91 KB)

Fig. 6. Model working on the MRF All-time stock data.

The historical analysis of MRF stock is illustrated through various graphical representations. Fig. 2 presents the Volume of MRF All-time data, highlighting the trading activity over the years. Fig. 3 depicts the Moving-Average of MRF All-time stock data, smoothing short-term fluctuations to reveal long-term trends. The Closing price of MRF data is shown in Fig. 4, capturing the overall price movement from 2004 to 2025. Fig. 5 illustrates the Test data of MRF All-time stock price, comparing the model's prediction against actual data. Finally, Fig. 6 demonstrates the Model working on the MRF All-time stock data, validating the model’s ability to learn and forecast stock trends based on historical performance. Challenges and Limitations in AI-Driven Stock Market Prediction

4.5 Data Quality and Availability

Stock market prediction relies heavily on data. Poor data quality, missing values, and noise in the data can significantly reduce the performance of predictive models.

4.6 Market Volatility and Unpredictability

Stock markets are influenced by various factors, including political events, natural disasters, and economic changes, which introduce high volatility. Predictive models often struggle to handle these unforeseen events.

4.7 Over-fitting and Model Generalization

AI models may over-fit to training data, making them ineffective on new data. Ensuring generalization and avoiding over-fitting requires careful tuning and often large datasets, which may not always be available.

4.8 Ethics and Regulation

The ethical and regulatory aspects of using AI in trading are increasingly under scrutiny. Issues such as market manipulation, lack of transparency, and potential job displacement call for stricter regulatory measures.

6 Conclusion

Stock market prediction using machine learning algorithms has been vastly improved over the past years. These kinds of models have been shown to potentially outperform traditional statistical methods by leveraging complex patterns in large datasets. These range from linear regression, to support vector machines (SVM), all the way to random forests, and even deep learning models like LSTM's and RNN's. Such models are able to capture the nonlinear relationships and detect trends, and can also forecast the short-run tendency more accurately. However, their performance relies on the quality of the data-set, feature construction and hyper-parameters.

However, due to the nature of the stock market (which is stochastic and unpredictable), this model can hardly be expected to accurately predict the stock market. ML models are very informative but being ML models, are not spared from macroeconomic turbulence, changes in regulation or once-in-a-lifetime occurrences like a pandemic. Recently, these highly efficient tools should to become an assistance to human expert not a decision-maker.

References

[1] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications, 42(1), 259–268.

[2] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669.

[3] Yu, P., & Yan, X. (2020). Stock price prediction based on deep learning models: A comprehensive review. Neural Computing and Applications, 32(7), 10765–10789.

[4] Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 83, 187–205.

[5] Shah, D., Isah, H., & Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. International Journal of Financial Studies, 7(2), 26.

[6] Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55(1– 2), 307–319.

[7] Kumar, M., & Thenmozhi, M. (2016). Forecasting stock index movement: A comparison of support vector machines and random forest. Decision, 43(1), 1–20.

[8] Huang, W., Nakamori, Y., & Li, Q. (2005). Support vector machine prediction of stock market movement. Expert Systems with Applications, 29(2), 251–259.

[9] Chen, C. L. P., & Chen, A. L. P. (2008). Applying support vector machines to financial time series forecasting. Expert Systems with Applications, 34(4), 1003–1009.

[10] Niu, X., Zhang, Y., Wang, H., & Zhang, J. (2020). Stock trend prediction: A deep learning approach. IEEE Transactions on Neural Networks and Learning Systems, 31(1), 3–16.

[11] Zhang, Q., Liu, L., & Si, L. (2016). Deep learning for stock prediction. Computational Intelligence and Neuroscience, 2016.

[12] Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. Expert Systems with Applications, 36(3), 5932–5941.
[13] Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications, 42(20), 7046–7056.
[14] Tsai, C. F., & Hsiao, Y. C. (2010). Combining multiple feature selection methods for stock market prediction. Expert Systems with Applications, 36(4), 4566–4573.
[15] Huang, Y., & Peng, J. (2020). Stock price prediction using a hybrid model with data preprocessing, machine learning, and feature engineering. Applied Soft Computing, 91, 106213.
[16] Zhong, X., & Enke, D. (2017). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. Financial Innovation, 3(1), 1–23.
[17] Wang, J., & Vergara, P. (2016). Sentiment-enhanced linear regression model for stock market forecasting. Decision Support Systems, 85, 77–88.
[18] Hsu, C. M. (2011). A hybrid procedure for stock price prediction by integrating self-organizing map and support vector regression. Expert Systems with Applications, 38(5), 6056–6062.
[19] Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Comparison of ARIMA and artificial neural networks models for stock price prediction. Journal of Applied Mathematics, 2014.
[20] Tay, F. E. H., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. Omega, 29(4), 309–317.
[21] Chong, T. T., & Ng, W. K. (2008). Technical analysis and the London stock exchange: Testing the MACD and RSI rules using the FT30. Applied Economics Letters, 15(14), 1111–1114.