ITM Web of Conferences **70**, 04017 (2025) https://doi.org/10.1051/itmconf/20257004017
DAI 2024

Forecasting the Treasury Yield Spread for FRED T10Y2Y Data Based on Multiple Approaches

Ying Chen Leu

American International School of Cape Town, Cape Town, South Africa

Abstract. As a matter of fact, price prediction for financial underlying assets always been a hot topic in financial fields in recent years with high volatility. With this in mind. this study looks into the usage of machine learning models to predict the yield spread between 10-year and 2-year US Treasury bonds (T10Y2Y). Based on the data from the Federal Reserve Economic Data (FRED) database (1976-2024), this study assesses the performance of four different models: multi-layer perceptron (MLP) regression, LSTM, ARIMA model and Facebook Prophet model. Each model's performance is measured using MAE, MSE, RMSE as well as F2 score. The results show that both MLP regression and LSTM models achieve high accuracy in predicting the yield spread. However, MLP regression outperforms LSTM in terms of producing more reasonable future predictions, particularly over longer time periods. ARIMA and Prophet, while effective for linear forecasts, were confused by the data and made unreasonable and incorrect predictions

1 Introduction

As a matter of fact, the stock market is an important part of the global economy, reflecting economic growth and decline. To make wise decisions investors rely on market forecasts. There are many indicators used to forecast market trends. One of the more common ones is the yield spread between the 10-year and 2-year U.S. Treasury bonds (T10Y2Y). This spread is often used to measure the overall economic status and predict possible recessions in the near future. The T10Y2Y spread is the difference in interest rates between two [1]. Usually when the spread is positive, it means that investors expect economic growth. The spread doesn't predict the actual growth rather what investors themselves predict. A spread, approaching zero, which historically indicates a recession [1]. Its movement is an accurate reflection of future economic conditions.

The T10Y2Y spread has become a focus point for economists due to its accuracy in its predictions. The usefulness of this indicator is the ability to show investor intention of the outlook for the economy. A positive yield curve is resulted from investors demanding higher rates on long-term bonds when they predict fast economic development. At times of economic instability, the yield on short-term bonds could increase more than long-term bonds, creating an inverted yield curve, which has been shown to occur before every U.S. recession

Corresponding author: yleu@aisct.org

in the last 50 years [2, 3]. The yield curve is a useful measurement for understanding market dynamics and directing financial decision-making because of its capacity to predict future economic decline [2].

In recent years, the increased complexity of markets has resulted to a growing interest in prediction methods like machine learning and deep learning. These prediction methods have been successful in stock market prediction, where they have become good at handling the vast amounts of data and capturing the non-linear patterns within datasets. Neural networks and support vector machines have demonstrated a great capacity to process and analyze big datasets, which makes them perfect for forecasting trends, prices, and other indicators such as the T10Y2Y in the stock market [4]. On the other hand, yield spread prediction using ML and DL is a relatively new technique. Time series data are typically predicted using forecasting techniques like ARIMA model. Developed through the Box-Jenkins methodology, ARIMA models are good for modeling linear relationships within the data [5]. Nevertheless, these models often struggle to predict when the data shows non-linear patterns, which are common in predicting yield spread. As a result, this study is going to explore how other, more advanced techniques might be used to improve yield spread forecasts.

Since deep learning algorithms can learn from large datasets and see patterns that others models would overlook. These models generally outperform simpler models. For example, LSTM networks, a type of neural network, have been accurate in time series forecasting because of their ability to remember long-term dependencies in data [6]. Because of those reasons, LSTM models have advantages over traditional techniques, especially in predicting stock market trends [7]. Similarly, another neural network model, the Multi-Layer Perceptron (MLP) regression model might also be accurate in financial forecasting because they can capture the non-linear relationships in data that simpler models might overlook [8]. Another tool for forecasting is Facebook's Prophet model. Designed to handle missing data and seasonal trends more effectively than many traditional methods [9], Prophet is particularly useful for business forecasting applications. Due to its adaptability and simplicity, financial time series forecasting is beginning to use it. Prophet is an essential tool for predicting in dynamic contexts because, which can swiftly adjust to changes in data patterns.

In a paper written in 2022, researchers proposed a novel scheme, integrating CNN, LSTM networks, and ARMA models. The study highlights the importance of accurate financial forecasting and addresses the limitations of traditional models like ARMA, which often fail to capture nonlinear patterns in financial data [10]. In another paper written in 2022, researchers explored the use of Facebook Prophet for predicting stock prices, demonstrating its potential for long-term financial forecasting. The study focuses on the importance of accurate stock price prediction, highlighting the challenges posed by the stock market's non-linear and unpredictable nature. The Efficient Market Hypothesis suggests that precise prediction of stock prices is nearly impossible, yet the study argues that with the right variables and models, reasonable accuracy can be achieved. The authors conclude that the FB-Prophet model provides better prediction accuracy and a lower error rate compared to traditional models like ARIMA. They suggest that combining FB-Prophet with other methodologies could further enhance performance and address scalability issues for larger datasets [11].

2 Methodology

The methodology for implementing various models to predict the yield spread between 2 Treasury bonds (T10Y2Y) involves a synchronized approach, beginning with data preprocessing and followed by model-specific steps, then to creating graphs of future predictions.

The LSTM model is designed to capture temporal dependencies. The first step involves preprocessing the historical T10Y2Y data by converting date values into datetime format and scaling the yield spread using MinMaxScaler. A key aspect of the LSTM model is defining a time step of 60 days. The data is reshaped into a 3D format required by the LSTM (`samples, time steps, features`) and split into training (80%) and testing (20%) sets. It is built using a Sequential architecture with two LSTM layers. The first LSTM layer, with 50 units, returns sequences to the second layer, which also has 50 units but does not return sequences. These are followed by two Dense layers, with the final layer outputting . The model is compiled using the Adam optimizer and MSE as the loss function. After training for 10 epochs with a batch size of 64, predictions are made. These predictions are then inverse-transformed to their original scale for evaluation. Finally, the model's predictive capability is assessed, including generating a forecast for the next 365 days.

The MLP (Multi-Layer Perceptron) model shares similar data preparation steps with the LSTM model. The MLP model uses a neural network structure with two hidden layers, each consisting of 100 neurons. This architecture is designed to capture non-linear data. The model is configured to run for a maximum of 500 iterations, with a random seed set to ensure the reproducibility of results. The same performance metrics used in the LSTM model, i.e., MAE, RMSE, and R^2, are applied here to assess the model's accuracy.

The ARIMA model is particularly effective that exhibit linear relationships and consistent patterns over time. After preprocessing the T10Y2Y data and dividing it into training and testing sets, the ARIMA model is constructed using the `auto_arima` function. This function automates the process of identifying the best-fitting model parameters by testing different combinations of ARIMA orders (p, d, q) and selecting the one that minimizes the AIC) [12]. Once the optimal ARIMA model is identified, it is fitted to the training data. The model makes it particularly useful for datasets where these patterns are prominent. However, its performance may be limited in this scenario as there's no linear pattern.

The methodology for implementing the Facebook Prophet model to predict the T10Y2Y involves leveraging Prophet's strengths in handling the data, particularly when trends and seasonality are key factors. The data is then formatted to meet Prophet's requirements, with the date column renamed to ds and the yield spread values to y. Any missing values are removed to ensure data integrity. The dataset is separated into training (80%) and testing (20%) sets. Prophet is then initialized, and the model is fitted to the training data. One preference of Prophet is its ability to automatically handle and model seasonality, trends, and even holidays, without requiring extensive manual parameter tuning. This makes it particularly efficient for quick deployment in forecasting tasks.

After fitting the model, Prophet generates a future dataframe that spans the entire testing period and beyond. This dataframe is used to produce predictions (yhat), which are then extracted from the forecast and compared against the actual values in both the training and testing datasets. Prophet's default settings assume certain seasonal and trend patterns, which are automatically detected and incorporated into the forecast. This automation is particularly useful that exhibits regular seasonal cycles or clear long-term trends. Prophet's design is well-suited for business forecasting applications where seasonality and trend components are predominant. However, it's important to recognize that while Prophet can be powerful, its default assumptions about underlying patterns may need adjustment or fine-tuning, depending on the specific characteristics of the data.

https://doi.org/10.1051/itmconf/20257004017

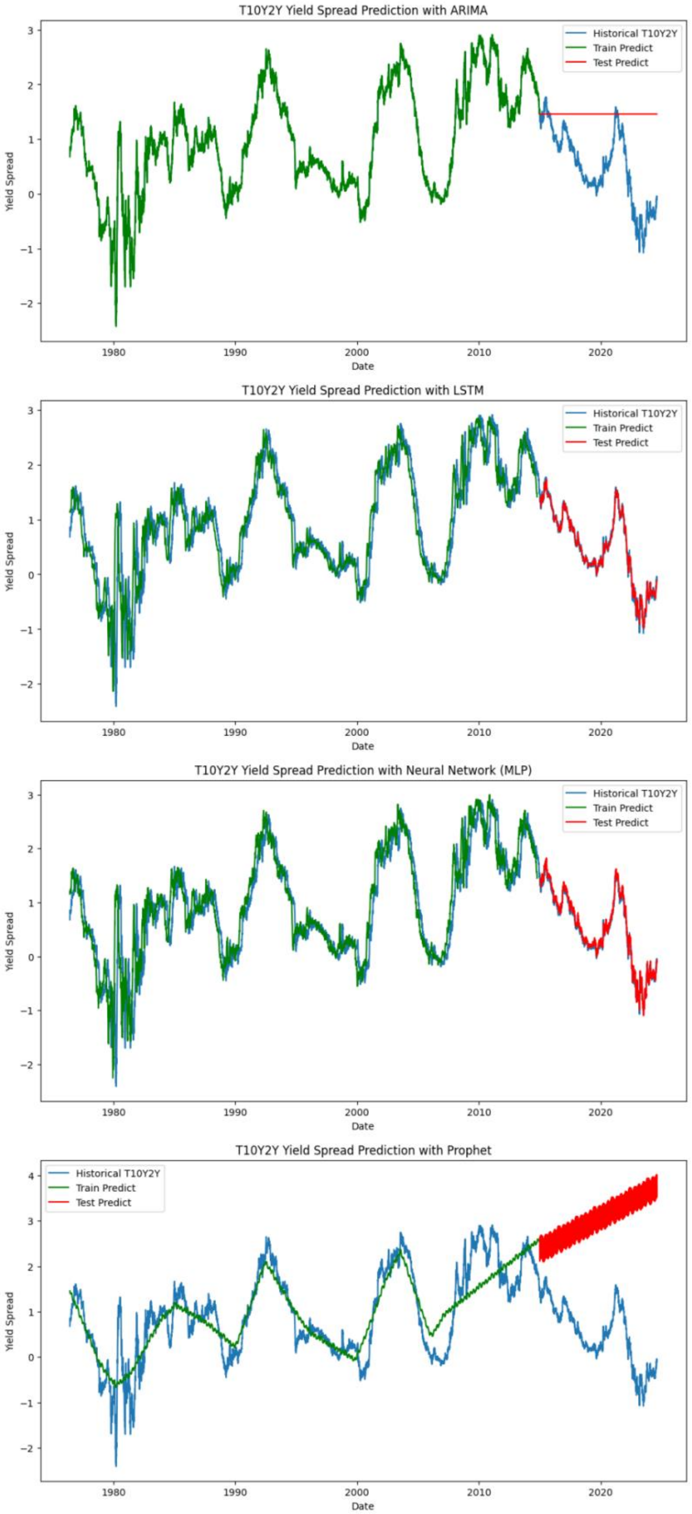

Fig. 1. Training and testing plot for different models (Photo/Picture credit: Original).

3 Results and discussion

The results for the models of training and testing sets are given in Table 1 and Table 2, respectively. Then, this study will analyze the performance of the four different models, i.e., LSTM, MLP, ARIMA, and Prophet, used to predict the T10Y2Y yield spread. Each model was evaluated based on their training and testing accuracy, and their future predictions were analyzed to assess their practicality and reasonableness. The training and testing plots are shown in Fig. 1 and the future trends are illustrated in Fig. 2.

Table 1. Training data accuracy metrics.

	MAE	MSE	RMSE	R Squared
LSTM	0.0570	0.0073	0.0857	0.9921
ARIMA	0.0013	0.0000	0.0019	1.0000
Prophet	0.3667	0.2292	0.4787	0.7506
MLP	0.7011	0.7570	0.8701	-22.1957

Table 2. Testing data accuracy metrics.

	MAE	MSE	RMSE	R Squared
LSTM	0.0439	0.0037	0.0606	0.9919
ARIMA	0.9716	1.3730	1.1717	-2.0340
Prophet	2.6513	8.0967	2.8455	-16.8921
MLP	0.4638	0.3081	0.5551	-18.3584

The LSTM model performed exceptionally well during both phases, with a training MAE of 0.0570, MSE of 0.0073, and an RMSE of 0.0857. The R-squared was 0.9921, indicating it captured most of the variance in the training data. Testing results were similarly strong, with a slightly better MAE of 0.0439 and RMSE of 0.0606, alongside an R-squared value of 0.9919, demonstrating the model's ability to generalize effectively. However, the future predictions made by the LSTM model reveal a significant issue. The model predicts a steady upward trend, resulting in an unrealistic and overly optimistic future forecast. The yield spread, according to the LSTM model, would recover from its negative values and approach zero by late 2025, which seems unreasonable given the complex economic factors influencing such trends. This suggests that while the LSTM model excelled at fitting historical data, it might be overfitting or overly influenced by recent trends when forecasting the future.

The Multi-Layer Perceptron (MLP) model underperformed compared to other models during the training and testing phases. The model recorded a high MAE of 0.7011 and an MSE of 0.7570 during training, leading to a poor RMSE of 0.8701 and a highly negative R-squared value of -22.1957. These poor results were echoed in the testing phase, where the model exhibited an MAE of 0.4638, an MSE of 0.3081, and an RMSE of 0.5551, with a similarly poor R-squared value of -18.3584. Despite these underwhelming performance metrics, the future predictions made by the MLP model were surprisingly the most reasonable among the four models. The MLP model forecasted a fluctuating yield spread, which aligns more closely with the volatile nature of financial markets. This suggests that while the MLP struggled with fitting the historical data, it was more cautious and realistic in its future predictions, avoiding the extreme projections seen in other models.

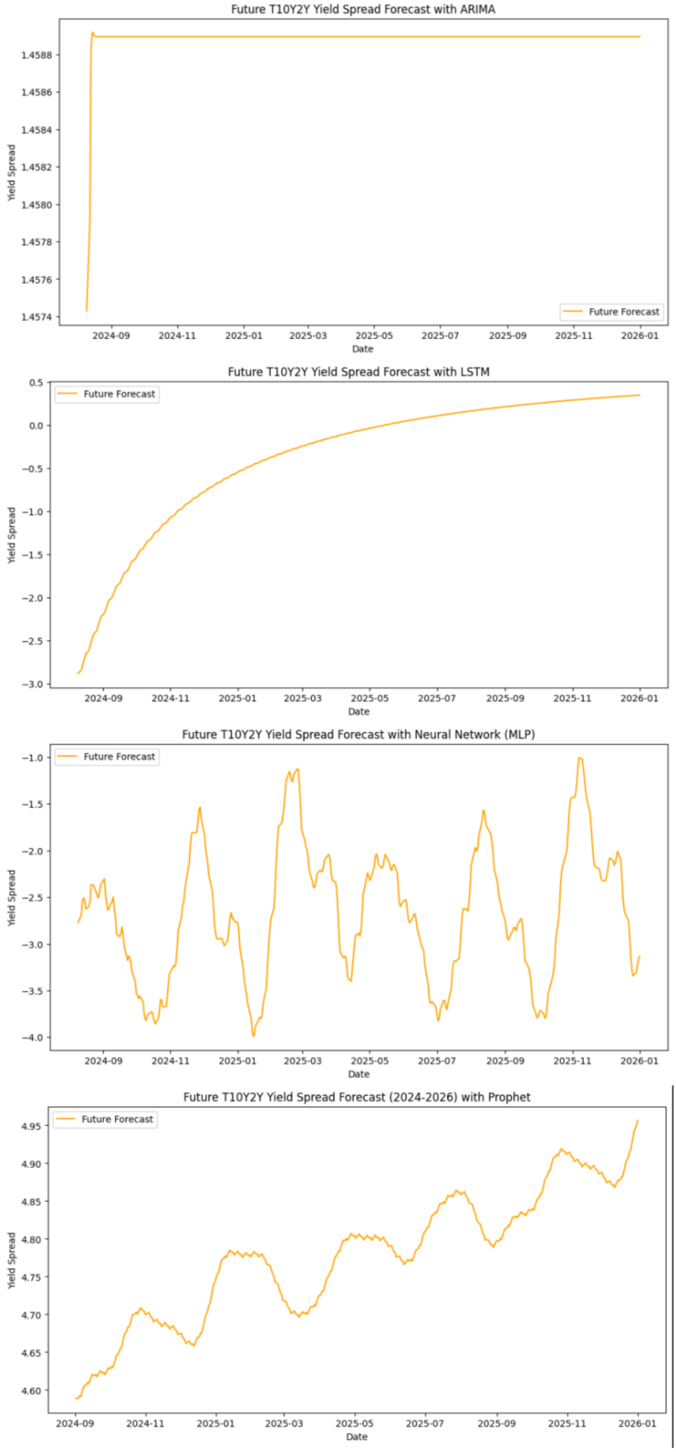

Fig. 2. Future prediction for different models (Photo/Picture credit: Original).

The ARIMA model displayed excellent performance during the training, with almost perfect results (MAE of 0.0013 and RMSE of 0.0019, R-squared of 1.0000). However, this performance did not translate to the testing phase, where the model recorded a very high MAE of 0.9716 and an RMSE of 1.1717, with a negative R-squared value of -2.0340. These results indicate that the model was overfitted to the training data and failed to generalize to new data. The future forecast made by the ARIMA model was particularly unrealistic. The model predicted a near-constant yield spread, hovering slightly above 1.45% throughout the forecast period. Such a flat forecast does not account for the expected volatility and economic factors that typically influence the yield spread, making the ARIMA model's prediction highly questionable.

The Facebook Prophet model, designed for handling trends and seasonality with minimal manual intervention, recorded modest training results with an MAE of 0.3667, MSE of 0.2292, and an RMSE of 0.4787. However, its testing performance was poor, with a MAE of 2.6513, MSE of 8.0967, and RMSE of 2.8455, accompanied by a highly negative R-squared value of -16.8921, indicating significant overfitting and poor generalization. The future forecast generated by Prophet was perhaps the most extreme of all models. Prophet predicted a rapid and significant increase in the yield spread, projecting values rising from 4.6% to nearly 5% by early 2026. This upward trend is unrealistic given the historical context and economic conditions, highlighting the model's limitations in this specific financial forecasting context.

While the LSTM model performed exceptionally well in terms of fitting the historical data and generalizing to the test set, its future forecast was overly optimistic, indicating potential issues with overfitting or trend assumptions. The ARIMA model, despite its near-perfect training performance, failed to generalize and produced an overly simplistic and unrealistic future prediction. The Prophet model, designed for scenarios with strong trends and seasonality, significantly overestimated future values, making it unsuitable for this specific task. Interestingly, despite its poor performance on historical data, the MLP model generated the most realistic future forecast, capturing the shifts of financial markets. This suggests that even when a model underperforms in traditional accuracy metrics, it may still provide valuable insights in terms of realistic predictions.

4 Conclusion

To sum up, this study explored the performance of four different models, i.e., LSTM, MLP, ARIMA, and Prophet, in predicting the T10Y2Y yield spread. Each model showed varying degrees of success, influenced by its underlying structure and the nature of the data. The LSTM model performed exceptionally well in fitting historical data, demonstrating high accuracy in both phases. This success is due to LSTM's ability to capture complex patterns. However, its future predictions were overly optimistic, suggesting that while LSTM can model trends within the training data effectively, it might overfit recent trends and fail to account for the inherent unpredictability of financial markets. The MLP model, although it struggled with fitting historical data, surprisingly provided the most reasonable future forecasts. MLP's simpler architecture doesn't capture time dependencies as effectively as LSTM, but this simplicity also means it avoids overfitting to recent data trends. As a result, its predictions were more conservative and realistic, reflecting the volatility typically seen in financial markets. The ARIMA model, which relies on linear assumptions, excelled in the training phase but failed to generalize to new data. Its nearly perfect training performance did not carry over to the testing phase, where it produced unrealistic and flat predictions. This indicates that while ARIMA is effective in capturing linear relationships, it struggles with the non-linear, complex nature of financial data. The Prophet model, designed to detect trends and seasonality automatically, did not perform well with the T10Y2Y yield spread. The model's assumptions about trends led to significant overestimation in future predictions, highlighting its limitations in dealing with financial time series that do not exhibit strong seasonal patterns. Prophet's tendency to overfit and extrapolate trends resulted in unrealistic forecasts, making it less suitable for this type of data.

There were several limitations in this study that affected the performance and generalization of the models. The dataset, while valuable, was limited in scope and may not have captured all the economic factors that influence the yield spread. Expanding the dataset to include a broader range of economic conditions could improve the models' accuracy and generalizability. Additionally, the models themselves have inherent limitations. For example, LSTM requires careful tuning to avoid overfitting, while ARIMA's reliance on linear patterns restricts its effectiveness in capturing the complexities of financial markets. Looking ahead, future research could explore hybrid models that combine the strengths of different approaches, such as integrating LSTM's capacity for capturing long-term dependencies with MLP's ability to produce realistic forecasts. Additionally, incorporating external economic indicators could enhance the models' context and predictive power. In summary, while this study demonstrated the different capabilities of different models in forecasting the T10Y2Y yield spread, it also emphasized the importance of refining these models to better address the challenges posed due to the intrinsic of financial markets. Continuous updates and improvements to these models will be essential for maintaining their relevance and accuracy in future predictions.

References

1. E.C. Engstrom, A. S. Sharpe, (Don't Fear) The Yield Curve, Reprise (Federal Serves, 2022).
2. M. D. Chinn, L. Ferrara, The Predictive Power of the Term Spread and Financial Variables for Economic Activity across Countries. National Bureau of Economic Research 32084 (2024)

   https://www.nber.org/system/files/working_papers/w32084/w32084.pdf.
3. J. Sabes, J.-G. Sahuc, The Predictive Power of the Term Spread for Recessions in the Euro Area. Empirical Economics, **64(1)**, 1-26 (2023)

   https://link.springer.com/article/10.1007/s42979-021-00592-x.
4. K. Sharma, R. Bhalla, G. Ganesan, Time Series Forecasting Using FB-Prophet. ACM **2022**, 59-65 (2022).

   https://www.scirp.org/journal/paperinformation?paperid=109491.
5. R. Hasse, L. Quentin, Predicting Recessions with the Yield Curve: A Bayesian Approach, Lund University (2022)
6. Deep Learning - Introduction to Long Short Term Memory. GeeksforGeeks, Retrieved from: https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/.
7. K. Sharma, R. Bhalla, Stock Price Prediction Using Multilayer Perceptron Neural Network. International Archives of Education and Technology, **2(1)**, 2021.

   https://ojs.sgsci.org/journals/iaet/article/download/162/152.
8. Stock Predictions using Multilayer Perceptron Regression GitHub, 2021, Retrieved from: https://github.com/Sinchana-SH/Stock-Predictions-using-Multilayer-Perceptron-Regression.
9. K. Sharma, M. Mahajan, A. Janghale, B. Joshi, Stock Price Prediction using Facebook Prophet. ITM Web of Conferences, **49** (2022).

https://doi.org/10.1051/itmconf/20224403060

10. K. He, Q. Yang, L. Ji, J. Pan, Y. Zou, Financial time series forecasting with the deep learning ensemble model. Mathematics, **11(4)**, 1054 (2023).
    https://www.mdpi.com/2227-7390/11/4/1054.
11. K. Sharma, R. Bhalla, G. Ganesan, Time Series Forecasting Using FB-Prophet. ACM, **2022**, 59-65 (2022).
    https://ceur-ws.org/Vol-3445/PAPER_07.pdf.
12. Estimation and Order Selection. Forecasting: Principles and Practice, 2nd ed., OTexts, 2020, Retrieved from: https://otexts.com/fpp2/arima-estimation.html.