Predicting US Dollar Movements Using the VIX and Machine Learning

Sahayu Raja

Received November 21, 2024
Accepted July 17, 2025
Electronic access August 30, 2025

The US Dollar is a staple of international trade, affecting a wide range of people worldwide; therefore, predicting its price is imperative. The USDX is an index that measures the currency price of 6 major currencies against the Dollar, and is used as the dependent variable in this study. The VIX is a measure of stock market volatility and is used as the independent variable in this study. This study aims to use the VIX to help predict the USDX price with the power of machine learning. Current literature indicates that the USDX and VIX share a positive relationship. This is because investors view the US dollar as a safe-haven asset, and purchase it when there is market uncertainty. This study uses various machine learning methods to explore this relationship. Linear Regression and Huber Regression are examples of linear regression-based models used. Several nonparametric methods were also used, like a Decision Tree Regressor, a Random Forest Regressor, a Gradient Boosting Regressor, and a Neural Network. The results from the study show that the VIX and USDX share a positive correlation, although the relationship is not a strong one. In addition, it also demonstrated that the VIX and USDX likely share a nonlinear relationship, as the nonparametric models performed exceptionally well with it.

Keywords: volatility, dollar, currency, machine learning, regression

1 Introduction

Financial markets determine the state of the economy, and fluctuations in these market prices can affect billions of people across the world. One such market is the market for the US Dollar, as it serves as the reserve currency of the world and is a pillar in international trade. Therefore, the dollar can be a large indicator of economic prosperity, and predicting its outcomes could help foresee economic recessions and expansions [1]. This project aims to achieve this by looking for correlations between the US dollar index and the volatility index of the S&P 500.

The specific US dollar index that is used in this project is the USDX, commonly referred to by its ticker symbol DXY. The USDX was created in 1973 by the United States Federal Reserve and is now managed by the ICE Data Indices. The index is composed of six major currencies: the Euro (weighted 57.6%), the Japanese Yen (weighted 13.6%), the British pound (weighted 11.9%), the Canadian Dollar (weighted 9.1%), the Swedish krona (weighted 4.2%), and the Swiss franc (weighted 3.6%). The USDX went through one major change in 1999 when the Euro replaced the German mark, the French franc, the Italian lira, the Dutch guilder, and the Belgian franc [2]. The index serves as an accurate way to assess the dollar's overall performance, as it is a holistic measure of its performance against several different currencies. In short, the index is immune to independent changes in exchange rate; therefore, it is a good measure for the US dollar and the best choice for this study.

The specific volatility index that is used in this project is the CBOE Volatility Index, commonly referred to by its ticker symbol VIX. The index predicts the volatility of the S&P 500 over the next 30 days. The VIX achieves this through a series of calculations that are based on the CBOE-traded standard SPX options. "It estimates the expected volatility of the S&P 500 Index by aggregating the weighted prices of multiple SPX puts and calls over a wide range of strike prices" [3]. The VIX and the S&P 500 share an inverse relation, which means that an increase in the S&P 500 would cause a decrease in the VIX, and vice versa. The VIX is an accurate measure of volatility, and it is highly studied; therefore, it is the best choice for this study.

Literature Review

The US dollar is classified as a safe-haven asset [4]. "A safe-haven asset is a financial instrument that is expected to retain, or even gain value during periods of economic downturn" [5]. Investors purchase these assets during times of uncertainty [6]. As a result, the US dollar shows a positive relationship with the volatility index: when the VIX rises, the dollar tends to rise; when the VIX falls, the dollar tends to decline [7].

Reputable economists, Michael Melvin and Mark Taylor, analyzed the behavior of major currencies during the 2008 global financial crisis and found that the US dollar exhibited strong safe-haven characteristics. As market volatility spiked, measured by the VIX, investors shifted capital into dollar-denominated assets, causing the USD to appreciate even when domestic US conditions were weak. Their findings suggest that increased equity market risk translates to higher demand for the USD, reinforcing the theory that the VIX and USDX are positively correlated[8].

Bollerslev, Tauchen, and Zhou studied how investors react to changes in market volatility and found that investors demand extra return when bearing additional risk. They showed that when expected volatility goes up, measured by the VIX, it affects how investors value different assets. Even though their research focused on stocks, it suggests that rising volatility could also impact currency markets, like the US dollar. This supports the idea that the VIX might help predict when investors move money into safer assets like the dollar[9].

These previous studies have explored the relationship between the US Dollar and market volatility, often confirming a positive correlation. However, most of these studies rely on traditional statistical methods such as correlation analysis, VAR models, or GARCH frameworks. For instance, a study from Sokhanvar & Hammoudeh investigated how safe haven assets, including the US Dollar, respond to stock market risk and volatility. It used traditional econometric models like GARCH to evaluate the interactions between volatility and asset prices. The researchers confirmed that the US Dollar behaves as a safe haven during high-volatility periods. The study did not employ any predictive machine learning frameworks[4]. Furthermore, a study from Aliu et al. examined the VIX–USDX relationship using VAR frameworks. They found that sometimes the VIX seems to affect the USDX, but other times it does not, suggesting that the relationship is complex and non-linear. This study also did not use any machine learning models[10].

These previous studies neglect modern machine learning approaches, like Random Forest, Neural Network, and Gradient Boosting, to predict USDX movements. Therefore, this study fills a clear gap by applying a range of ML models. Furthermore, few of these studies use the VIX as a sole predictor, filling another research gap.

Understanding and predicting movements in the USDX is critically important for central banks, investors, and businesses, all of whom are impacted by currency fluctuations. This study offers a tool for improving such predictions by using the VIX, a symbol of market fear, as the sole predictor. Unlike prior research that relies solely on time-series methods like VAR or GARCH, this approach also employs a variety of different methods to correlate market volatility and currency prices. The findings can be used to demonstrate how investor sentiment alone may be sufficient to predict dollar trends. Furthermore, the study can highlight how modern machine learning can outperform traditional tools in modeling financial data.

The current research on the relationship between the US dollar and the VIX mostly supports the hypothesis that they share a positive correlation. This research paper aims to back up this theory using the power of machine learning. Several predictive models are used to identify price changes in the VIX and predict the change in the US dollar.

Methods

Dataset and Data Processing

The datasets used in this project are the CBOE Volatility Index and the USDX data. Both datasets originate from Yahoo Finance. The VIX dataset is linked here, and the USDX dataset is linked here. The data for this study was sourced from Yahoo Finance due to its credibility, accessibility, and wide historical coverage. While small discrepancies exist in the data, it was cleaned for null values, and general trends were double-checked. Other data sources, like FRED, do not have the same level of accessibility and historical range as Yahoo Finance; therefore, Yahoo Finance was chosen for this study. All of the data is numerical data and is stored in a CSV file. There are seven columns in the data: Date, Open, High, Low, Close, Adj Close, and Volume. For the VIX data, the Date represents the days between 1990-01-02 and 2024-06-21. For the Dollar data, the Date represents the days between 1971-01-04 and 2024-07-15. However, some dates in these periods are missing, and some dates hold NaN values. The Open column represents the price of the index at the start of trading. The High represents the highest point the index reached in the trading window, while the Low represents the lowest point the index reached in the trading window. The Close column represents the price of the index at the end of the trading window. The Adj close column represents the close amount after adjustments for splits and dividends. But this column's data is the same as the close column in our dataset. The Volume data is 0.0 for all the rows in both datasets, as indexes cannot be traded; therefore, no data displays their volume.

The original dollar dataset had 16600 rows with 7 columns, and the original VIX dataset had 8994 rows with 7 columns. After filtering out the null values, the USDX dataset had 13597 rows with 7 columns, and the VIX dataset had 8684 rows with 7 columns. This means that there were 3003 null values for the USDX dataset and 310 null values for the VIX dataset. Removing missing values helped avoid bias because it stopped the model from learning from incomplete or inaccurate data. Incomplete data can confuse the model, making it harder to learn real relationships between the VIX and USDX. If we had filled in the blanks with guesses, it could have created false patterns and made the results less reliable. Therefore, removing the null values was the best decision. There were no duplicate entries in both datasets; therefore, the number of rows remained unchanged after the function was performed. The Volume and Adj Close columns were dropped, as they were unnecessary and repetitive, bringing the dataset down to 5 columns. As mentioned above, the USDX dataset and the VIX dataset have different dates. The inconsistency in these dates made it extremely hard to conduct any data analysis; therefore, the two datasets were combined into one dataset. The new dataset has 8682 rows with 9 columns. The new columns are: Date, Vix Open, Vix High, Vix Low, Vix Close, USDX Open, USDX High, USDX Low, and USDX Close. Once the new dataset was compiled, two new columns were added to it: USDX Change and Vix Change. These columns measured the day-to-day price change of the indexes. The first row in both columns was 0.0, as there was no prior date to refer to. This addition brought the total number of columns up to 11. Another dataset was created with the lagged differences of the VIX. The dataset had three columns: VIX Close, Yesterday Close, and Day Before Close. This dataset is used in a multivariable regression model. To split the datasets into testing and training datasets, sci-kit learn's train_test_split method was used. The method randomly splits the datasets, ensuring that the model is trained on a wide range of data. 80% of the dataset was used for the training dataset, while 20% of the dataset was used for the testing dataset.

Machine Learning Models

Several machine learning methods were used to conduct this study, ranging from simple regression models to complicated decision trees and neural networks. All methods were implemented with the help of the scikit-learn Python library. The first method that was tried was a linear regression model. This model aimed to create a line of best fit between the VIX and the USDX. This line would then be used to predict USDX. However, the dataset had a lot of outliers, and the model was severely impacted by them, so it would have been ideal to use a loss function that is less sensitive to outlier data. As a result, several outlier-resistant regression models were tried. Huber loss regression, Quantile regression, and RANSAC regression were all used. These methods employ different methods to discourage outlier data from affecting the line of best fit. Another tactic used to nullify the effect of outliers was to limit the range of VIX values in the training set and test sets. As a result, a linear regression model was constructed using all observations in the set where the VIX was less than 30. The last type of regression model that was tried was multivariable regression. This method employs the use of multiple x variables to predict the price of the USDX. The variables that were used were the VIX price, yesterday's VIX price, and the day before's VIX price, apart from these regression models. Several nonparametric models were used. A decision tree regressor, a neural network regressor, a gradient boosting regressor, and a random forest regressor were all tried in this study. The decision tree regressor was chosen because it is easy to understand and can handle non-linear relationships without needing a lot of data preparation. The neural network regressor was chosen because neural networks are powerful at detecting complex patterns and were used to see if deeper trends exist between the VIX and the dollar index. Gradient boosting was included because it builds strong predictions by combining many small models and often performs well on structured data like this. The random forest regressor reduces overfitting by averaging multiple decision trees and is good for handling noisy or unpredictable financial data. The decision tree regressor had an overfitting problem, in which it performed well on the training data, but not as well on the testing data. The overfitting ratio was around 0.6044, clearly illustrating its superior performance on the training data. With the help of sklearn's GridSearchCV method, the models were able to have their parameters hyper-tuned, solving the overfitting problem. After the change was made, the overfitting ratio moved to around 0.9701, a value that is extremely close to 1, showing the balanced performance of the model across the training and testing data.

Results

The table below shows the mean absolute error for every model. It is split into testing and training columns. The mean absolute error measures the absolute difference between the predicted value and the actual value.

Table 1 shows the results from the machine learning models. The models that performed the best were the Decision Tree Regressor, the Gradient Boosting Regressor, and the Random Forest Regressor. The models that performed the worst were the Quantile Regression model and the RANSAC Regression model.

The table 2 shows some of the linear regression-based models and their coefficients.

In Table 2, the First 30 Linear Regression model had the largest slope, followed by the Huber Regression model, the Linear Regression model, and the Multiple Variables Linear Regression. All of the coefficients are positive.

Discussion

The results from this research are promising. The mean absolute error ranged from about 7.5 to 8.2, meaning that the USDX can be predicted to a satisfactory level using the various models. The original hypothesis, which was that the USDX and the VIX share a positive correlation, holds to some extent. All of the linear models had a positive coefficient, but their coefficients were small. The coefficients were never above 0.5. This indicates that there is a positive correlation, but it isn't a strong one. The idea that the two indices are positively correlated is consistent with most other studies in the field. The p-values for the Linear Regression Model and Huber Regression model are ¡0.001, indicating that the relationship between the VIX and the

Table 1 Mean Absolute Error on training and testing data for each machine learning model. All values are rounded to the thousandths.

Method Name	Mean Absolute Error Testing	Mean Absolute Error Training
Linear Regression	8.127	8.012
Huber Regression	8.078	7.994
Quantile Regression	8.280	8.113
RANCAS Regression	8.121	8.075
Linear Regression on First 30 VIX Values	7.928	7.695
Multiple Variable Linear Regression	8.109	8.012
Decision Tree Regressor	7.800	7.567
Neural Network Regressor	8.122	8.009
Gradient Boosting Regressor	7.799	7.469
Random Forest Regressor	7.757	7.523

Table 2 Coefficients of the linear regression-based models. The values are rounded to the nearest thousandth.

Method Name	Coefficient(s) / P-values
Linear Regression	0.159 / 0.000
Huber Regression	0.253
First 30 Linear Regression	0.416 / 0.000
Multiple Variables Linear Regression	0.035, 0.060, 0.070 / 0.637, 0.540, 0.344

USDX is statistically significant. This suggests that changes in the VIX have a strong impact on the USDX. However, the p-values for Multiple Variables Linear Regression were above 0.05, indicating that none of the predictors were statistically significant. This means that this model does not provide strong evidence that the VIX can reliably predict the USDX. The nonparametric models also had promising results, as they performed better than the linear models in most cases. Three of the best-performing models were all nonparametric models, indicating that there is a relationship between the VIX and the USDX, but that it might not be linear. The variation in model performance can be attributed to the nature of the relationship between the VIX and the USDX. Linear models assume a constant relationship, which does not capture the complexity of the volatility market. This is a problem in finance because relationships between things like prices and volatility change over time, and big price swings often happen close together. Simple models aren't good at handling that kind of behavior. As shown by their low coefficients and high mean absolute errors, these models were limited in predictive power. In contrast, nonparametric models outperformed these linear approaches. This suggests that the VIX–USDX relationship is not strictly linear and follows a non-linear relationship. Many other studies have explored the relationship between the VIX and the USDX; however, none of them are directly comparable to this study, as they do not employ the same machine learning models. These studies concluded that the VIX and the USDX have a slight positive correlation, but there is a stronger nonlinear relationship at play.

The results from this research can have various applications around the world. It can be used to predict the US dollar price, helping government institutions and businesses plan day-to-day operations. The study is limited, however, in that it only has a few independent variables, which are the VIX and its lagged iterations. Some next steps in the study would be to include other safe-haven assets as independent variables, like gold, bonds, and bitcoin. Another way to add to this study would be to study independent currency pairs, instead of an aggregate measure like the USDX.

Conclusion

This study can serve as a stepping stone for other researchers in the field. The study concluded that the USDX and VIX share a mild positive correlation. Since the nonparametric models performed exceptionally well, the study also concluded that the two indices may share a relationship that is not linear. One limitation of this study is that it only uses the VIX as an independent variable, which excludes other important factors that influence the USDX. Variables such as interest rates, inflation, and commodity prices can have a significant impact on currency movements, and their exclusion limits the model's overall accuracy. Other researchers could build off this study by adding different independent variables and currency pairs to create a stronger model. The study's findings are significant because they demonstrate that the VIX can be a meaningful predictor of the USDX when analyzed using machine learning models. This highlights the value of applying modern modeling techniques to financial data. This research contributes to the growing field of machine learning–based financial forecasting and provides a foundation for future studies to build from. Furthermore, the study has various implications for the real world. Businesses that partake in international trade, and whose business is vulnerable to changes in exchange rates, could use one of the models to help run operations. Forex traders who specialize in trading the U.S. Dollar could use the model to help execute trades. Government institutions, like the Federal Reserve, could use these models to aid in creating monetary policy and tools. Even everyday people could use the model to plan investments. Models like these are critical to the lives of everybody in modern society. The financial system is an intricate and complex web of different financial instruments that are connected. Exploring the relationship between these instruments is critical to understanding the financial system as a whole. As more studies are made about different financial instruments, the financial market becomes clearer and more predictable, helping businesses, governments, retail investors, and society as a whole.

References

1 C. Bertaut, B. Beschwitz and S. Curcuru, *The international role of the U.S. dollar: Post-COVID edition [Internet*, https://www.federalreserve.gov/econres/notes/feds-notes/the-international-role-of-the-us-dollar-post-covid-edition, Available from:.

2 J. Chen, *U.S. dollar index – USDX definition [Internet*, https://www.investopedia.com/terms/u/usdx.asp, Available from:.

3 J. Kuepper, *CBOE volatility index (VIX) definition [Internet*, https://www.investopedia.com/terms/v/vix.asp, Available from:.

4 A. Sokhanvar and S. Hammoudeh, *Comparative analysis of responses of risky and safe haven assets to stock market risk before and after the yield curve inversions in the U.S [Internet*, https://www.sciencedirect.com/science/article/abs/pii/S105905602400368X, cited 2025 Jun 9]. Available from:.

5 I. Analyst, *What are safe-haven assets and how do you trade them?*, Available from:https://www.ig.com/en/trading-strategies/what-are-safe-haven-assets-.

6 J. Chai, C. Zhao, Y. Hu and Z. Zhang, *Structural analysis and forecast of gold price returns*.

7 *Correlation between the VIX index and the U.S*.

8 M. Melvin and M. Taylor, *The crisis in the foreign exchange market*.

9 T. Bollerslev, G. Tauchen and H. Zhou, *Expected stock returns and variance risk premia*.

10 F. Aliu, A. Asllani and S. Hašková, *The impact of bitcoin on gold, the volatility index (VIX), and dollar index (USDX): analysis based on VAR, SVAR, and wavelet coherence*.