Joseph Damico
02/02/2025
Applied Data Science
DSC680-T303 (2253-1)

Final Report: Forex Market Volatility Prediction

Business Problem

The Forex market is the largest and most liquid financial market, but it is also highly volatile and unpredictable. Traders rely on technical and fundamental analysis to make decisions, but market movements often follow a random walk pattern making accurate predictions challenging.

This project aimed to predict Forex market volatility using historical OHLC (Open, High, Low, Close) data for the **EUR/USD** currency pair. The goal was to assess whether machine learning and statistical models could effectively capture price movement trends and provide actionable insights for traders.

Background/History

The Forex market operates 24 hours a day with global participants trading currencies for speculation, hedging, and international trade. Volatility in this market is influenced by macroeconomic factors, geopolitical events, and central bank policies. However, previous studies suggest that price movements in Forex tend to be random walks making short-term predictions difficult.

Understanding volatility can be beneficial for risk management, trade timing, and position sizing, even if absolute price direction remains unpredictable.

Joseph Damico
02/02/2025
Applied Data Science
DSC680-T303 (2253-1)

Data Explanation

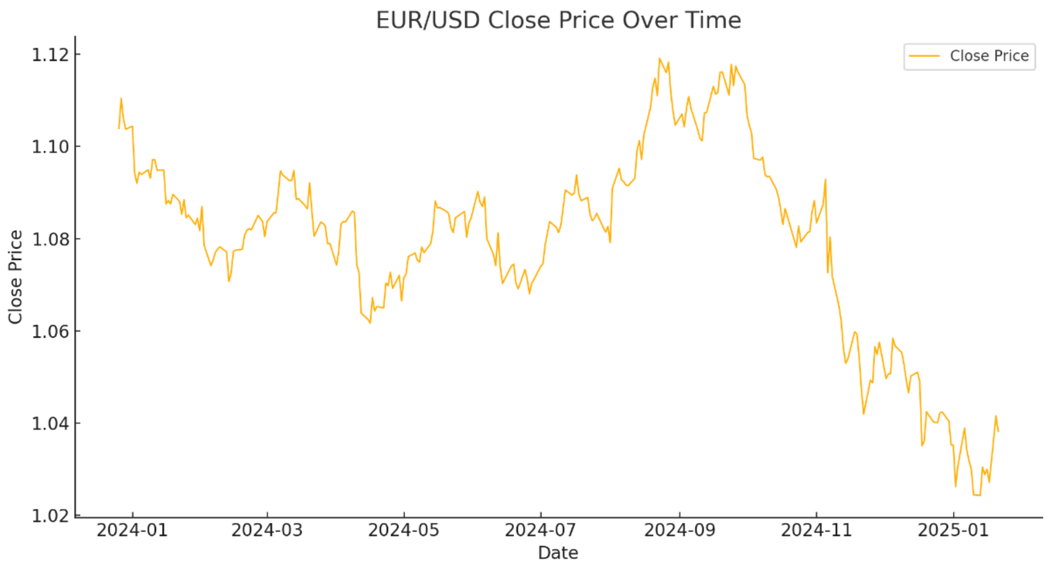
Data Source

- **TradingView**: The dataset was sourced from TradingView, containing historical OHLC price data for the **EUR/USD** currency pair in a daily timeframe.

Data Dictionary

Feature Name	Description
• **Date**	Timestamp of each observation
• **Open**	Opening price of the trading day
• **High**	Highest price reached during the trading day
• **Low**	Lowest price reached during the trading day
• **Close**	Closing price of the trading day
• **Daily Returns**	Percentage change in the closing price
• **ATR (Average True Range)**	Volatility indicator measuring price range movements
• **Bollinger Bands**	Upper and lower price bands based on standard deviations from a moving average

The following visualization highlights the historical **EUR/USD** close price trends over the dataset period.



Methods

The project used three predictive approaches to model Forex market volatility:

Random Forest Regressor

- Supervised machine learning model leveraging Daily Returns, ATR, and Bollinger Bands.
- Evaluated using Root Mean Squared Error (RMSE) and R^2 Score.

ARIMA (5,1,0)

- A traditional time-series forecasting model that captures autoregressive patterns in price movements.
- Focused on forecasting the Close price based on historical values.

Proposed LSTM

- A neural network-based sequential model designed to capture long-term dependencies in time-series data.
- Due to limitations in my current ability implementation was proposed but not executed.

Analysis

Random Forest Results

- **RMSE**: **0.0287**
- **R^2**: **-2.73** (very poor correlation between features and target)

Interpretation:

The model failed to predict market movements effectively due to the lack of strong correlations between historical data and future price action.

ARIMA Results

- **RMSE**: **0.0175**
- **R^2**: **-0.39** (slightly better but still poor)

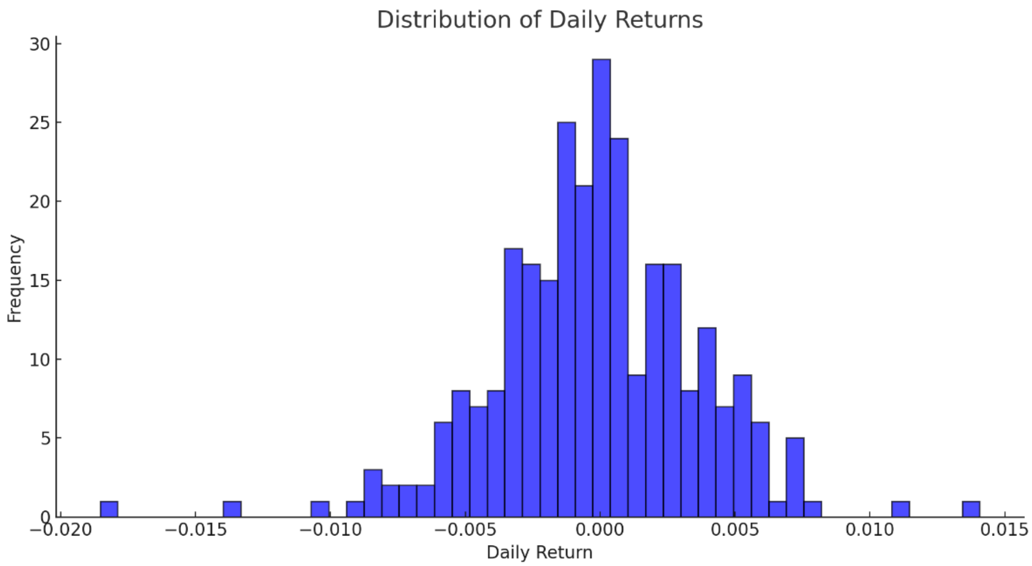
Interpretation:

ARIMA performed better than Random Forest but was still ineffective further proving the random nature of Forex price movements.

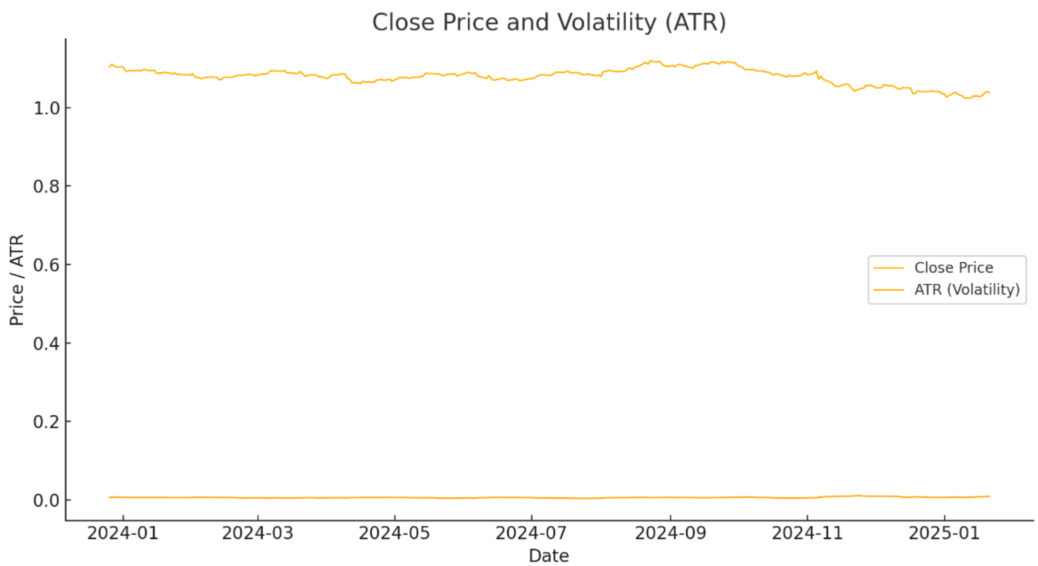
Joseph Damico
02/02/2025
Applied Data Science
DSC680-T303 (2253-1)

Volatility Indicators

This histogram shows that most daily returns cluster around zero, meaning large price changes are rare and unpredictable.



The ATR indicator captures bursts of market volatility demonstrating periods when price swings were more extreme.



Conclusion

- Forex prices exhibit random walk behavior making short-term predictions unreliable.
- Traditional models struggled to capture market trends with low correlation between features and target values.
- The LSTM model could potentially offer improved results by capturing sequential dependencies in price action.

Assumptions

- Historical price data is a fair representation of market conditions.
- ATR and Bollinger Bands provide meaningful insights into volatility.

- Feature selection was appropriate for modeling price movements.

Limitations

- Macroeconomic factors (e.g., interest rates, news events) were not included.
- The dataset was limited to **EUR/USD** and may not generalize to other currency pairs.
- The LSTM model was not implemented.

Challenges

- Random walk behavior limits predictability.
- Low correlation between engineered features and future prices.
- Limitations prevented LSTM testing in this environment.

Recommendations

- Use LSTM models for deeper learning of Forex trends.
- Combine fundamental and technical indicators for better predictive power.
- Integrate results into a trader’s decision-support system rather than relying solely on predictions.

Q&A Section

1. What motivated you to focus on Forex volatility prediction?

Joseph Damico
02/02/2025
Applied Data Science
DSC680-T303 (2253-1)

Forex markets are among the most liquid and volatile in the world, yet they exhibit strong elements of randomness. I wanted to explore whether machine learning and time-series models could provide meaningful insights into price movements. Even though the results confirmed that Forex follows a random walk, the process of testing different models helped in understanding the challenges of financial forecasting.

2. Why did you use historical OHLC data instead of including macroeconomic factors?

The initial goal was to evaluate whether technical analysis indicators alone could provide reliable signals for price prediction. While macroeconomic factors such as interest rates and GDP growth influence Forex markets, incorporating them would require external data sources and introduce complexity in feature engineering.

3. How accurate were the models, and how can they be improved?

- The **Random Forest model** had an **RMSE of 0.0287** and an **R^2 of -2.73**, indicating a poor fit.
- The **ARIMA model** improved slightly, with an **RMSE of 0.0175** and an **R^2 of -0.39**, but it still underperformed.
- **Future improvements**: Incorporating macroeconomic indicators and using deep learning models like LSTM could enhance prediction accuracy.

4. What were the biggest challenges faced in this project?

- Forex markets exhibit random walk behavior, making short-term price movements difficult to predict.
- Low correlation between features and future price movements reduced the effectiveness of machine learning models.
- Due to limitations, I was unable to test the LSTM model.

5. Why did you choose Random Forest and ARIMA instead of other models?

- Random Forest was chosen because of its robustness in handling complex patterns and feature importance analysis.
- ARIMA was tested as a classic time-series forecasting model that captures autoregressive and moving average components.
- LSTMs were considered as a future step because they are designed for sequential dependencies in time-series data.

6. What insights were gained from Exploratory Data Analysis (EDA)?

- Forex price movements showed limited correlation with engineered features.
- ATR (Average True Range) highlighted periods of increased volatility, which could be useful for identifying trading opportunities.
- Bollinger Bands provided a useful visual representation of price expansion and contraction but were not strong predictors for future price direction.

7. Why does ATR appear useful, but the models failed to predict volatility?

Joseph Damico
02/02/2025
Applied Data Science
DSC680-T303 (2253-1)

ATR successfully measures historical volatility but does not predict future volatility. Since volatility itself is unpredictable in a random market, the models struggled to find meaningful patterns.

8. How can this analysis be applied in real-world trading?

While the models struggled with precise predictions, the volatility indicators like ATR and Bollinger Bands remain useful for risk management and trade strategy development. Traders can use these features to adjust position sizing rather than trying to predict exact price movements.

9. What limitations did the dataset have?

- Excluded macroeconomic factors such as central bank policies, inflation rates, and geopolitical events.
- Limited to **EUR/USD** and may not generalize to other currency pairs.
- Did not incorporate real-time trading volumes, which could impact market behavior.

10. What are the next steps for improving this project?

- Implement LSTM neural networks for improved sequential learning.
- Integrate macroeconomic indicators and news sentiment analysis.
- Develop a real-time dashboard for Forex volatility monitoring.