Macroeconomic Drivers and Forecasting of the USD/CAD Exchange Rate

Chia-Yu Wei

2025-04-23

Abstract

This study investigates the macroeconomic drivers influencing the USD/CAD exchange rate and develops a hybrid forecasting framework that integrates both statistical and machine learning methods. Using monthly data from 2005 to 2023, the research explores how key economic indicators—such as interest rate differentials, inflation, GDP, and oil prices—affect the exchange rate between the U.S. and Canadian dollars. A Seasonal ARIMA (SARIMA) model is used to capture the linear and seasonal patterns in the exchange rate, while residuals from the SARIMA model are further modeled using machine learning algorithms (Random Forest and XGBoost) to account for nonlinear influences. The hybrid approach demonstrates improved predictive performance over individual models. This framework contributes to more accurate forecasting and informed decision-making in finance and policy contexts.

Background

The USD/CAD exchange rate is a critical economic indicator that reflects the relative strength of the U.S. and Canadian economies. It affects trade, investment, monetary policy decisions, and global financial markets. Given the high volume of economic activity and trade between the two countries, understanding the factors that drive changes in this exchange rate is critical. Traditionally, exchange rate modeling has relied heavily on time series techniques; however, these methods may overlook nonlinear relationships embedded in economic data. With the increasing availability of high-quality macroeconomic data and advances in machine learning, there is a growing opportunity to integrate diverse modeling techniques for improved forecasting accuracy. This study aims to explore this opportunity by examining the relationship between macroeconomic fundamentals and the USD/CAD exchange rate and proposing a hybrid model that leverages the strengths of both statistical and machine learning approaches.

Object of the Study

The primary objective of this study is to analyze how key macroeconomic indicators influence the USD/CAD exchange rate and to identify the most significant economic drivers behind its fluctuations. Specifically, the study aims to: (1) assess the impact of variables such as interest rate differentials, inflation differentials, GDP differentials, and oil price spreads on the exchange rate; (2) build a time series model that captures the underlying trends and seasonality in the data; and (3) enhance the forecasting accuracy by incorporating machine learning techniques to model the nonlinear components captured in the residuals. By developing a robust and interpretable hybrid forecasting framework, this research seeks to improve our understanding of exchange rate dynamics and provide valuable tools for economic forecasting and decision-making.

Data Description

Dependent Variables: USD/CAD monthly exchange rate

Independent Variables (US & Canada):

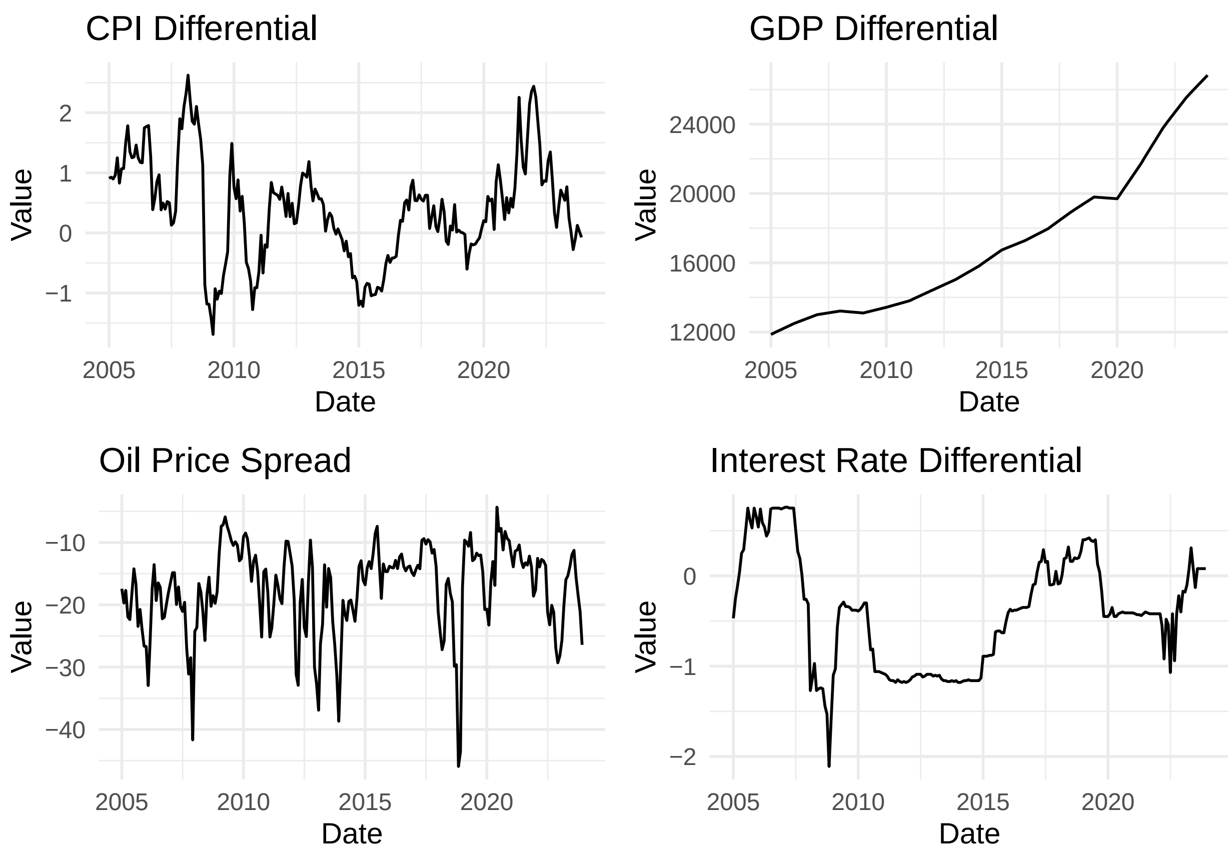
1. Interest rate differential
2. Inflation (CPI) differential
3. GDP differential
4. Oil price spread

Data scources and time frame (2005 ~ 2023)

Exploratory Data Analysis

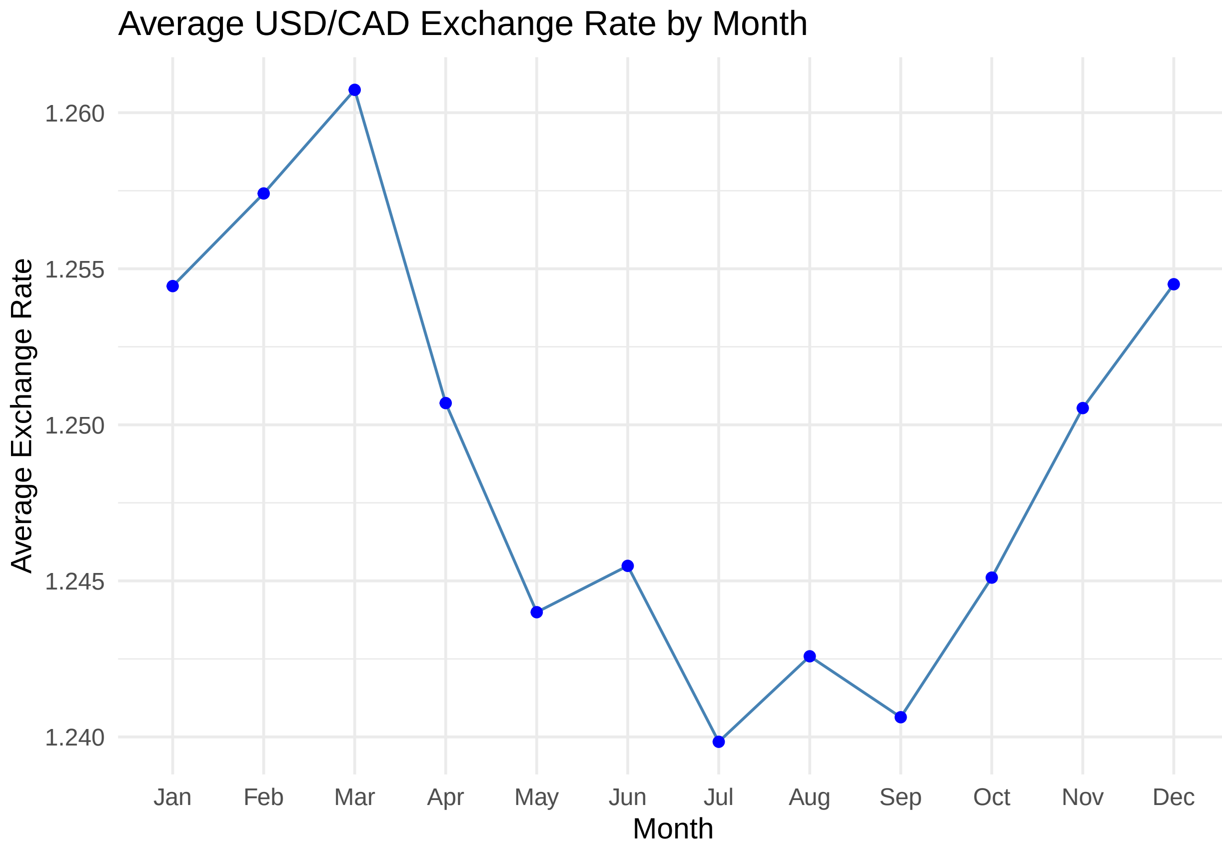
Time series plots of key economic variables

The time series plots of key variables—interest rate differential, inflation differential (CPI), GDP differential, and oil price spread—reveal distinct trends and fluctuations over the period from 2005 to 2023. Notably, the 2008 global financial crisis and the COVID-19 pandemic (2020–2022) coincide with significant shifts in several indicators, such as sharp drops in oil price spread and interest rates. The GDP differential exhibits a steady upward trend, while inflation and interest rate differentials show cyclical volatility, reflecting changing monetary policy between the U.S. and Canada. The oil price spread demonstrates high variance, with occasional extreme dips, likely reflecting global oil market shocks.



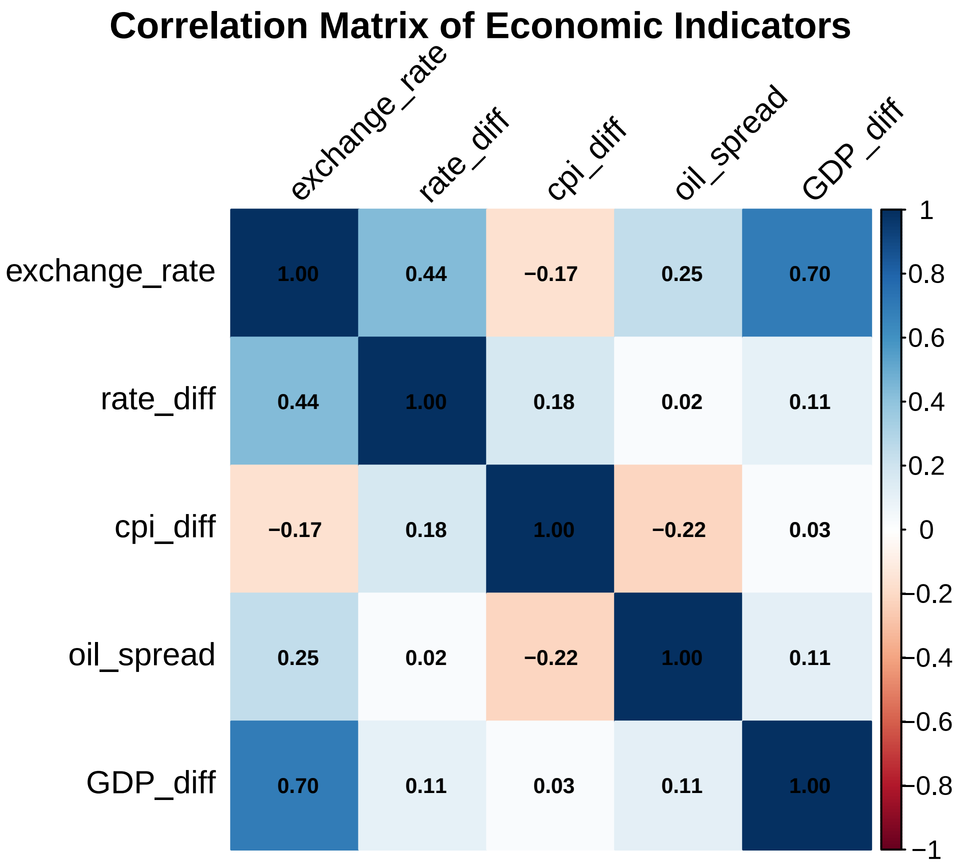
Average Exchange Rate by Calendar Month

The following figure shows the average USD/CAD exchange rate for each calendar month, aggregated across the full sample period (2005–2023). The exchange rate tends to peak around March and decline during mid-year months (July–September), suggesting a possible seasonal pattern. This visualization highlights temporal fluctuations that may be relevant when designing time series models or policy analysis



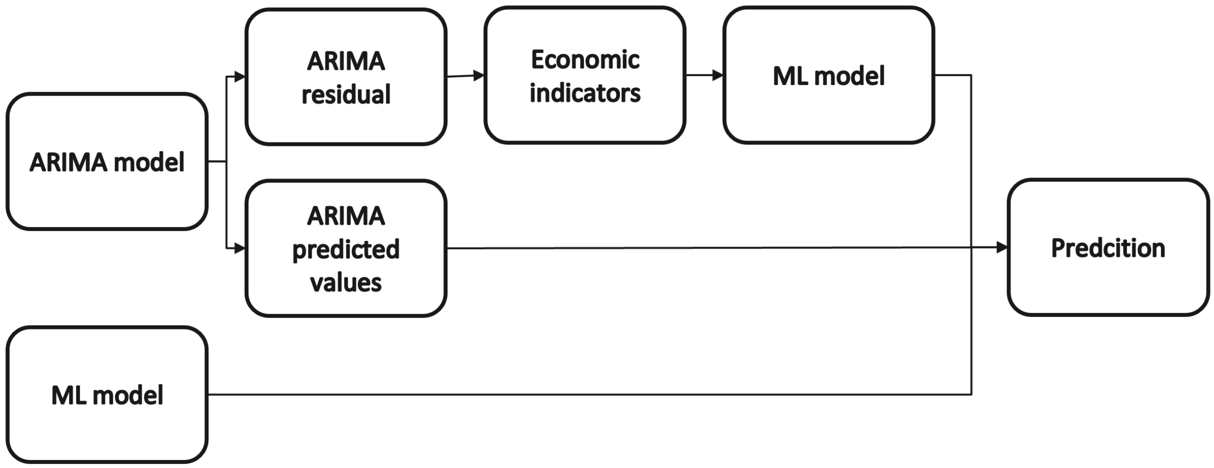
Correlation Matrix

Figure displays the correlation matrix among the key economic variables: exchange_rate, rate_diff, cpi_diff, oil_spread, and GDP_diff. The exchange rate exhibits the strongest positive correlation with GDP differential (0.70), suggesting that as the economic output gap between the U.S. and Canada widens, the USD tends to strengthen relative to the CAD. Interest rate differential also shows a moderate positive correlation with the exchange rate (0.44), indicating its influence on capital flows and currency valuation. In contrast, CPI differential has a weak negative correlation (-0.17) with the exchange rate, implying a less direct relationship. Oil price spread and rate differential appear to have little correlation with each other or with other variables. Overall, this matrix helps identify which variables may contribute most significantly to exchange rate fluctuations and supports feature selection for modeling.



Modeling Framework

During this study, I am to predict the future exchange rate 1 year from now. I came up with a hybrid modeling framework to forecast the USD/CAD exchange rate by leveraging both statistical and machine learning techniques. The framework begins with a Seasonal ARIMA (SARIMA) model to capture linear and seasonal patterns in the time series data. After fitting the SARIMA model, we extract the residuals—which represent unexplained variance, and model them using machine learning algorithms, including Random Forest and XGBoost. These models incorporate lagged macroeconomic indicators such as interest rate differential, inflation differential, GDP differential, and oil price spread as input features. The final forecast is obtained by summing the SARIMA prediction and the predicted residuals from the machine learning model. This two-stage approach combines the strength of SARIMA in modeling temporal dynamics with the flexibility of machine learning in capturing nonlinear effects, and I wanted to see if this method can usefully improve predictive performance. The figure below illustrates the proposed modeling framework.

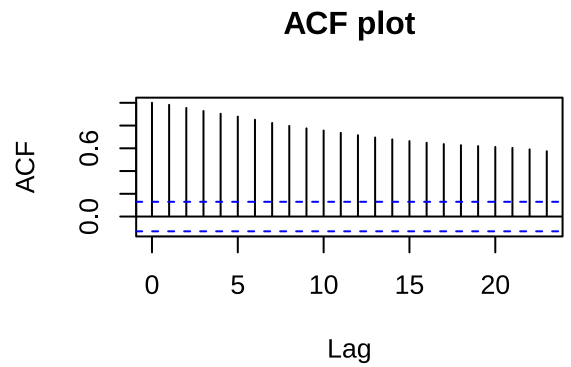


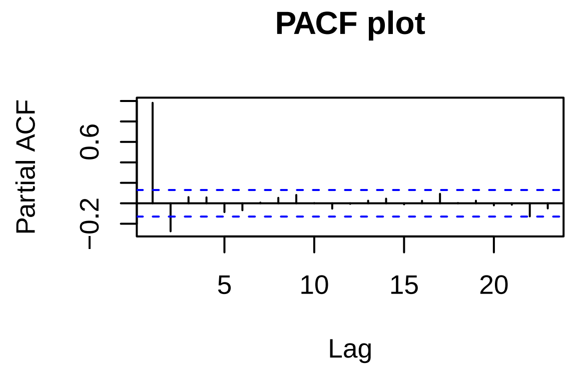
Time Series Analysis

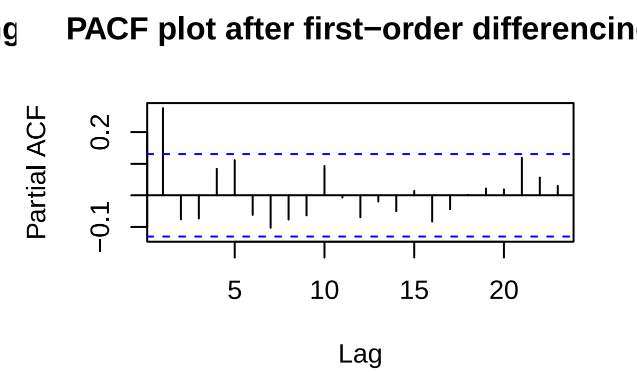
ACF plot

I plotted the autocorrelation graph to check the series' stationarity. The steady lowering of spikes confirms that the time series is nonstationary. Then I tried first-order difference of exchange rate, the ACF and PACF plots indicates the data is stationary. Therefore, the result supports me to use:

$$ARIMA(p, 1, q)$$







Seasonal Decomposition

The dataset is decomposed into trend, seasonal, and residual components in the time series decomposition. I used type = "addictive" here because I assume that the the time series is the sum of the components:

$$Observed = Tread + Seasonality + Residual$$

The result verifies the seasonal ARIMA (SARIMA) model using the US Dollar dataset's seasonality component.

1. Original Series: This is the original USD/CAD exchange rate data.

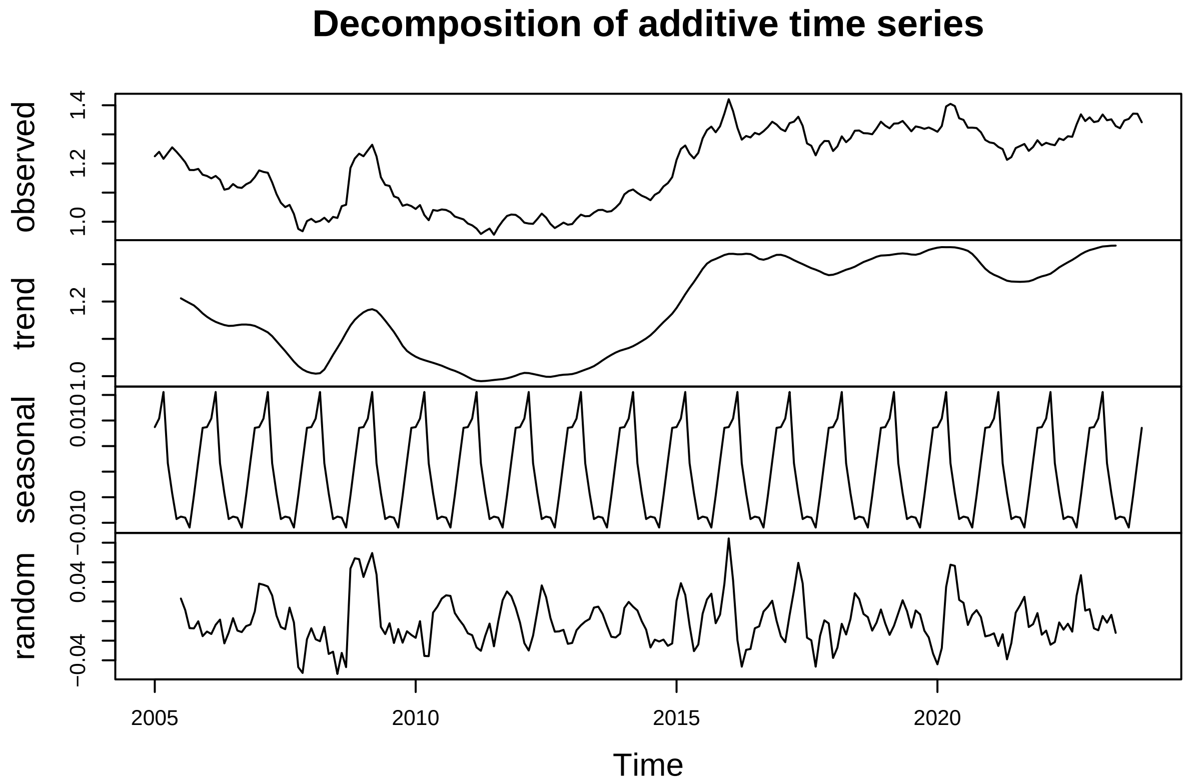
2. Trend Component: The trend line shows the underlying direction of the exchange rate after removing seasonal and irregular noise.

3. Season Component: The shape is repetitive every year, and the amplitude is consistent, meaning seasonality is present but stable over time.

4. Random/Residual Component: This is what's left over after removing the trend and seasonal parts. It appears to fluctuate around zero, with no strong patterns, indicating the decomposition was effective.

Therefore, I considered using SARIMA model:

$$SARIMA(p, 1, r)(P, 1, Q)_{12}$$



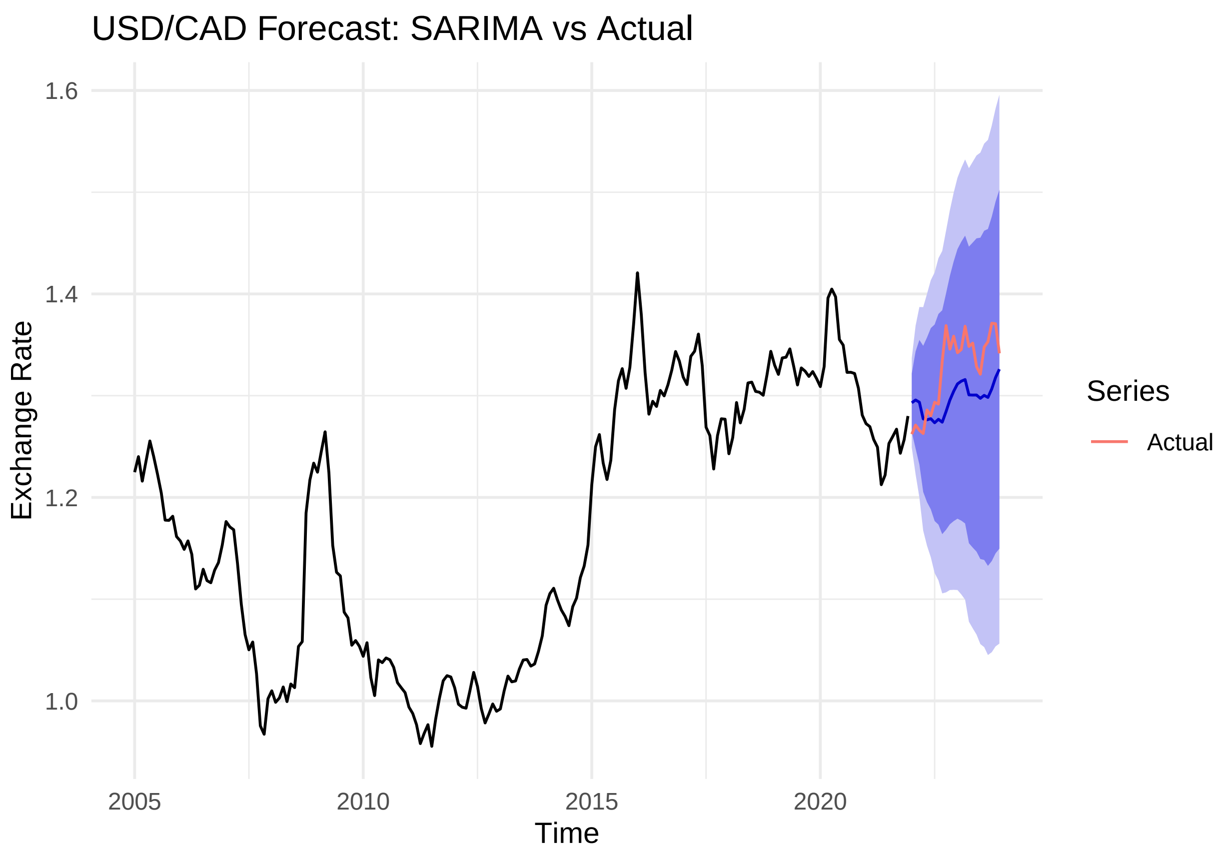
Model Implementation

I systematically tested multiple combinations of the hyperparameters (p, q, P, Q). For each candidate model, we computed the Akaike Information Criterion (AIC) and root mean squared error (RMSE) on the forecasted values to evaluate model performance. A lower RMSE indicates better out-of-sample predictive accuracy, while AIC helps penalize model complexity. The table below summarizes the performance of various parameter combinations. The model with (p=2, q=0, P=1, Q=2) achieved the lowest RMSE, suggesting it best balances model fit and forecasting accuracy among the tested options.

p	q	P	Q	AIC	RMSE
2	0	1	2	-874.5333	0.04222
1	2	1	2	-876.7646	0.04318
1	1	1	2	-873.7589	0.04339
1	0	1	2	-874.8324	0.04619
1	0	0	0	-764.4586	0.05022
0	2	0	0	-764.6880	0.05066

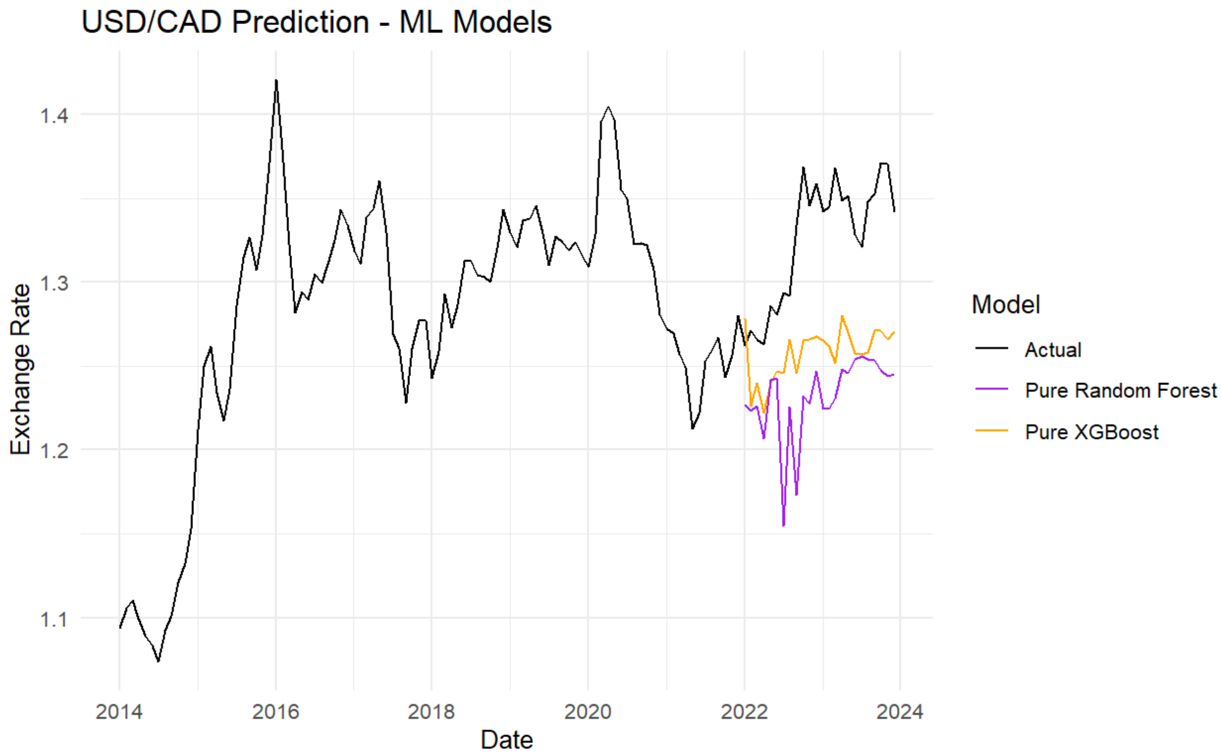
Forecasting using best SARIMA model

The SARIMA$(1,1,1)(1,1,0)_{12}$ model achieved the lowest RMSE. The plot below using the latest two years data to depict predicted values along with the test data. We can see the actual data stay within confidence bound, which support that the model capture the seasonality and trend well.



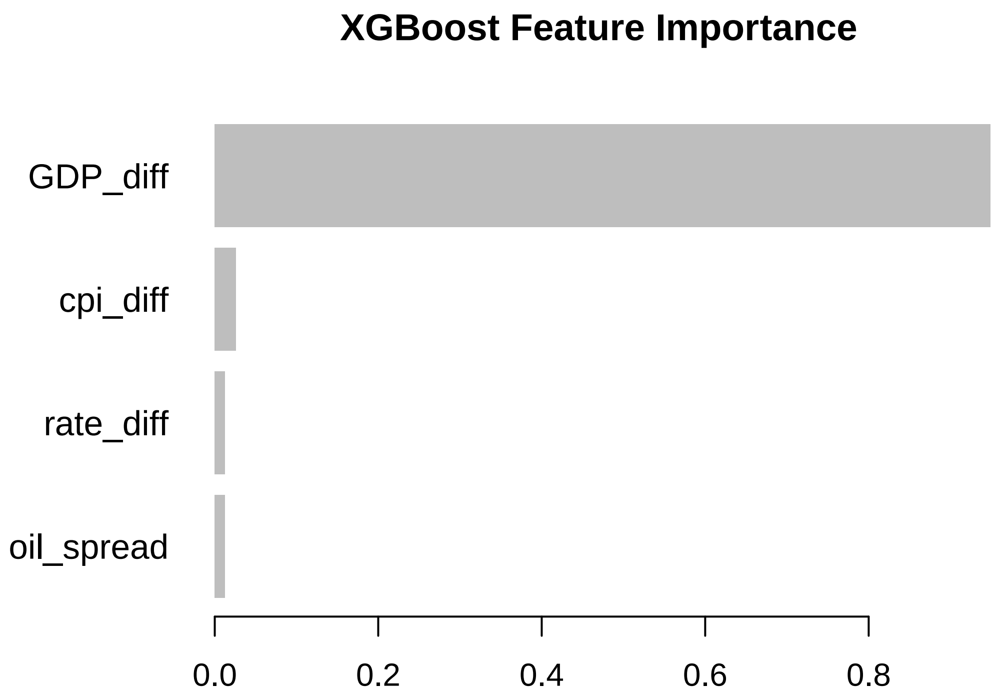
Machine Learning Model

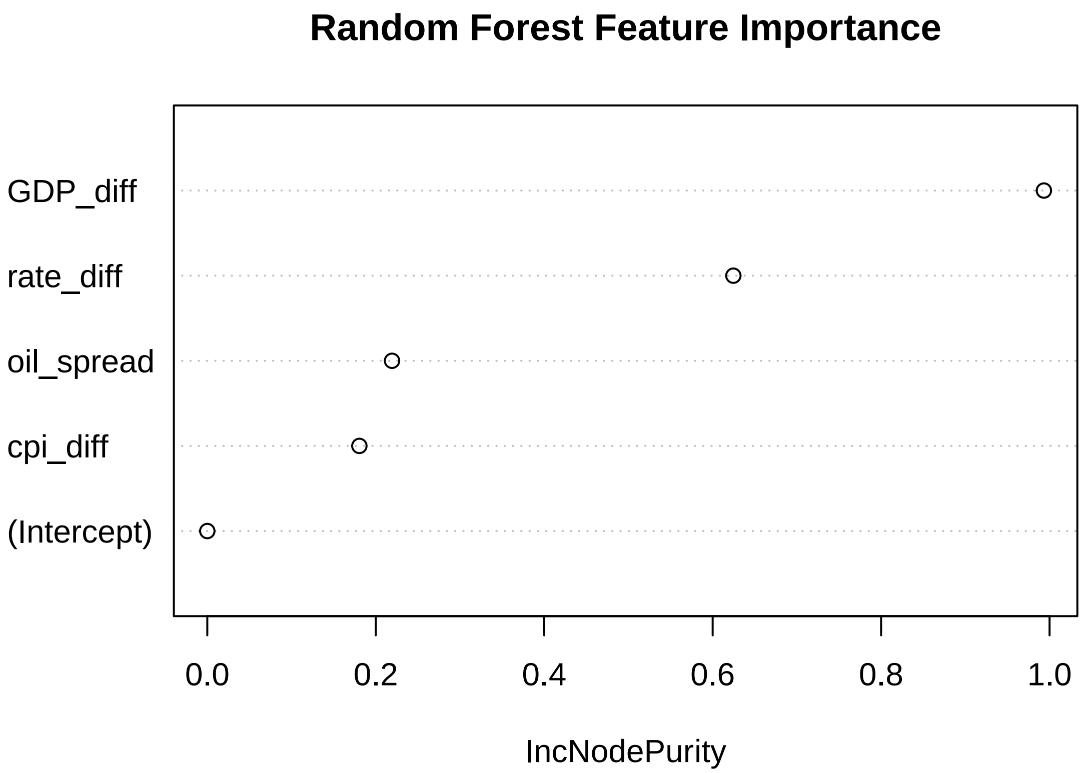
In addition to traditional time series modeling using SARIMA, we also explored the use of standalone machine learning models to predict the USD/CAD exchange rate. The objective was to assess whether non-linear methods, such as Random Forest and XGBoost, could outperform SARIMA in terms of predictive accuracy when used independently. By comparing the performance of both approaches as individual models, we aimed to evaluate the relative strengths of statistical versus machine learning techniques in capturing the underlying dynamics of the exchange rate.
The Figure shows the testing result, which illustrates the forecasting performance of two standalone machine learning models—Random Forest and XGBoost—on the USD/CAD exchange rate. Compared to the actual exchange rate trajectory, the XGBoost model captures a more stable pattern and stays relatively close to the average level of the actual exchange rate during the test period. In contrast, the Random Forest model underestimates the exchange rate throughout the forecasting horizon and exhibits limited responsiveness to fluctuations. This may be due to its lower extrapolation ability or sensitivity to data distribution shifts. Overall, while machine learning models can learn complex nonlinear relationships, their standalone performance appears to be less accurate in capturing the temporal dynamics of exchange rate movement compared to time series approaches. These results highlight the importance of integrating time-aware modeling techniques when forecasting financial time series.



Feature Improtance

The feature importance plots presents the feature importance rankings derived from the XGBoost and Random Forest models, both of which were trained to predict the USD/CAD exchange rate using macroeconomic indicators. In both models, GDP differential emerges as the most influential variable by a significant margin, indicating that differences in economic output between the U.S. and Canada play a dominant role in explaining exchange rate fluctuations. In the XGBoost model, GDP differential alone accounts for nearly the entire predictive power, while the contributions of other features—such as inflation differential, interest rate differential, and oil price spread—are minimal. The Random Forest model similarly highlights GDP differential as the most important variable, but also attributes moderate importance to interest rate differential and oil price spread, suggesting a more distributed feature influence. These results underscore the critical role of GDP gap in determining exchange rate movement and suggest that while machine learning models can handle multiple predictors, their effectiveness in this case is primarily driven by a single strong economic signal.


XGBoost Feature Importance
GDP_diff
cpi_diff
rate_diff
oil_spread
0.0
0.2
0.4
0.6
0.8

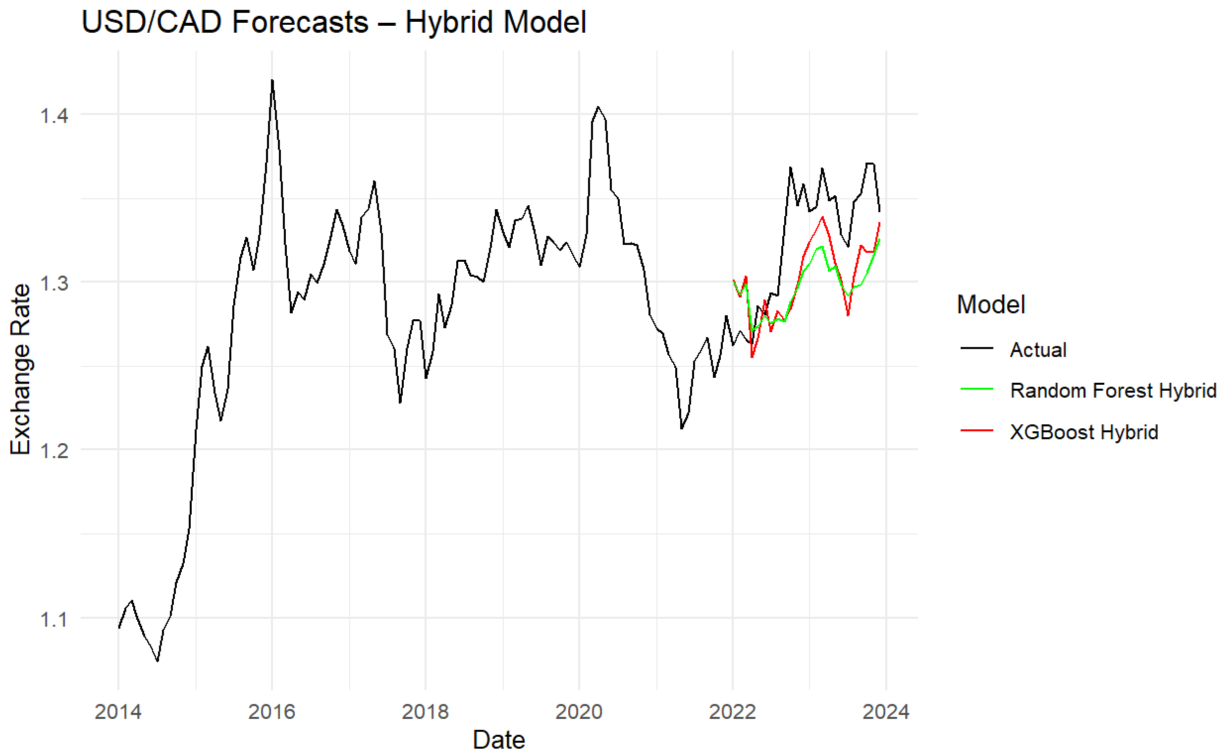


Hybrid Modeling

The hybrid model is implemented in two stages: first, a SARIMA model is used to capture the linear and seasonal structure of the exchange rate series. Then, the residuals from the SARIMA model are modeled using machine learning algorithms, allowing us to account for remaining nonlinear patterns.

Based on the testing result, we can see the hybrid model-XGBoost has best performance with lowest RMSE. Also, the testing plot of two hybrid model produce forecasts that closely follow the overall movement and directional trends of the actual USD/CAD exchange rate. The predicted lines successfully capture the general trajectory, turning points, and seasonal fluctuations observed in the real series, particularly during volatile periods. Although minor deviations in amplitude exist, the shape of the predicted curves aligns well with the actual values, indicating that the hybrid approach is effective at modeling both trend and short-term variations in the exchange rate.

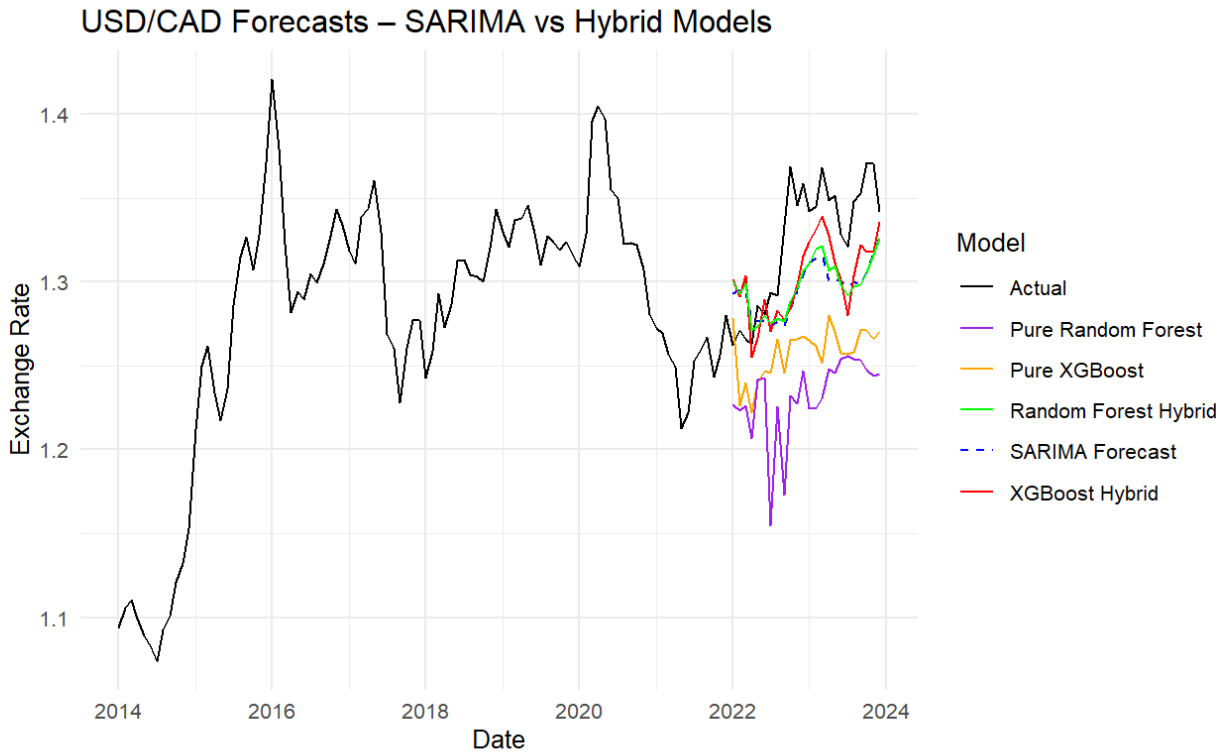
Model	RMSE
SARIMA	0.04222
Hybrid Random Forest	0.04394
Hybrid XGBoost	0.04123



Evaluation

As shown in Figure below, the SARIMA and hybrid models significantly outperform the pure machine learning models in terms of forecasting accuracy. Both pure Random Forest and pure XGBoost models exhibit relatively high RMSE values (0.11619 and 0.07589, respectively), indicating that relying solely on macroeconomic variables without time-dependent structure leads to suboptimal results. In contrast, the hybrid models—particularly the XGBoost Hybrid—achieve the lowest RMSE (0.04123), followed closely by SARIMA (0.04222) and Random Forest Hybrid (0.04394). The RMSE values among these top three models are very close, differing by less than 0.002, suggesting comparable predictive performance. However, given the small numerical differences.

Model	RMSE
SARIMA	0.04222
Random Forest	0.11619
XGBoost	0.07589
Hybrid Random Forest	0.04394
Hybrid XGBoost	0.04123



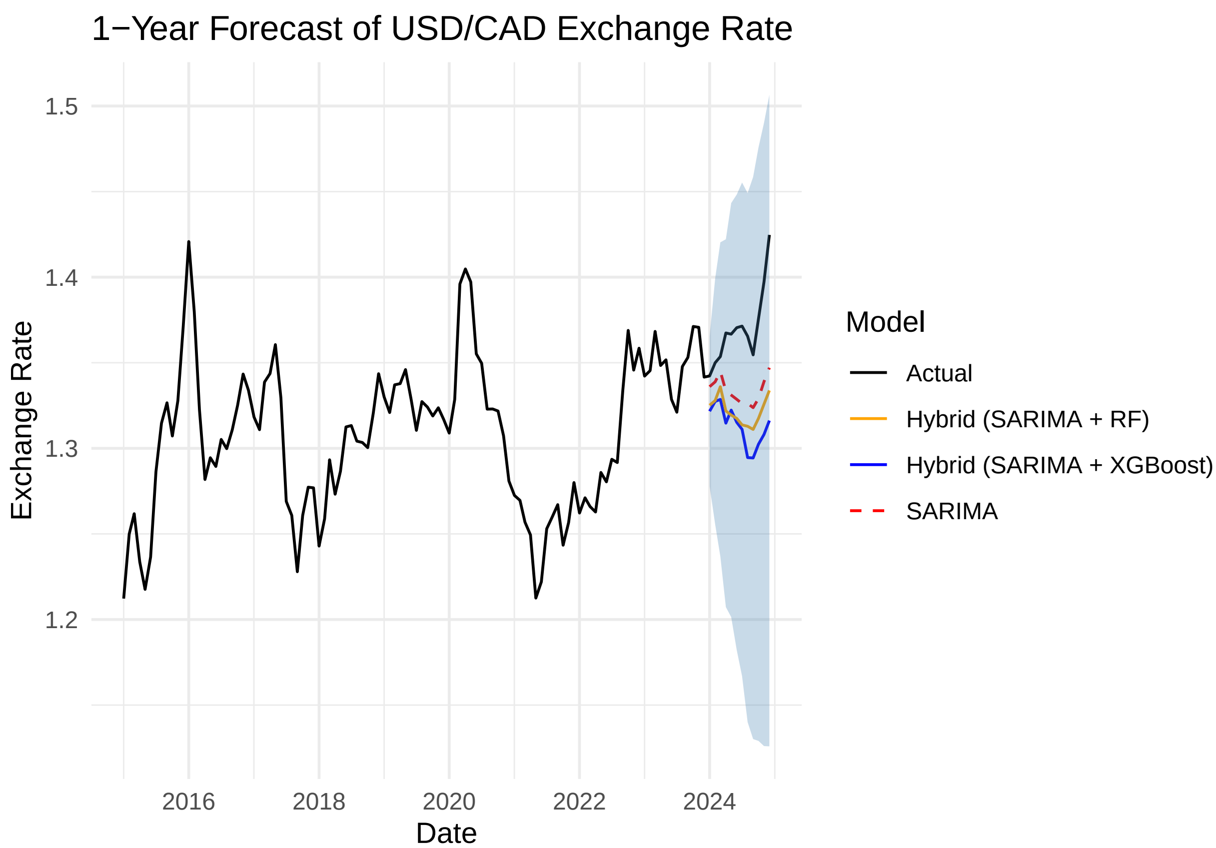
Forecast

In order to compare model performance under realistic forecasting conditions, we conducted a 12-step-ahead forecast, equivalent to predicting the USD/CAD exchange rate one year into the future. During model training, all models had access to historical exchange rate and macroeconomic data. However, in the actual forecasting phase, future macroeconomic variables were unknown. To address this, we first applied SARIMA and machine learning models separately to forecast each macroeconomic variable using their own past (lagged) values. These predicted values were then fed into the hybrid models as inputs for one-year-ahead exchange rate prediction.

Interestingly, although the hybrid models performed well during evaluation with observed (true) data, their predictive performance declined when using forecasted macroeconomic inputs—particularly when compared to the SARIMA-only model. Among all models, SARIMA achieved the best performance, suggesting that it not only provides stable predictions but may also have effectively captured the key underlying dynamics in the data.

A possible reason for the hybrid model's underperformance lies in the accumulated uncertainty from forecasting multiple input variables. Inaccurate predictions of macroeconomic features could have introduced compounded errors, limiting the hybrid model's ability to generate accurate final forecasts. This result highlights the sensitivity of hybrid approaches to input quality and reinforces the value of SARIMA's robustness in long-horizon forecasting scenarios. m

Model	RMSE
SARIMA	0.0414
Hybrid Random Forest	0.0516
Hybrid XGBoost	0.0624



Conclusion

In this study, we investigated the forecasting of the USD/CAD exchange rate using a combination of time series and machine learning approaches. We began with a SARIMA model to capture temporal trends and seasonal patterns, followed by standalone machine learning models and hybrid frameworks that integrated macroeconomic indicators. Our evaluation showed that while pure machine learning models underperformed, the hybrid models—particularly those using XGBoost—demonstrated promising accuracy when provided with actual economic data.

However, under realistic forecasting conditions where future macroeconomic variables must also be predicted, the SARIMA model consistently outperformed both hybrid and machine learning models. This suggests that SARIMA not only delivers robust and stable forecasts but also effectively captures key underlying signals in the exchange rate dynamics. The decline in hybrid model performance may be attributed to error propagation from the first-stage forecasting of economic inputs.

Overall, our findings highlight the trade-off between model complexity and input reliability. While hybrid models hold potential when accurate auxiliary forecasts are available, time series models like SARIMA remain a dependable choice for long-term exchange rate forecasting in scenarios with high input uncertainty.

Data Resource

- USD/CAD Exchange Rate: The USD/CAD exchange rate.
- Fed Rate: The U.S. Federal Funds Target Rate (%), set by the Federal Reserve. Higher values usually strengthen the USD.

- Canadian Interest Rate: The Bank of Canada's overnight policy interest rate (%). Like Fed_rate, it reflects central bank monetary policy, affecting CAD strength.
- US Consumer Price Index (CPI): The U.S. Consumer Price Index (CPI). It's an index value (not a percentage), representing average price levels.
- Canadian Consumer Price Index (CPI): Similar to US Consumer Index, but for Canada.
- WTI/WCS Oil Price