



Implementation of the ARIMA-GARCH Model on USD/JPY Exchange Rate Forecasting

Evelyn Zhu

Hunter College High School, New York, 10128, United States
evelynzhu@hunterschools.org

Abstract. Exchange rates exhibit unique market behavior with influences on many macroeconomic factors, making it an interesting subject for time series analysis. The Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models have been employed to address the topic, however, the liquidity and non-linearity of the foreign exchange market as a financial time series makes exchange rate forecasting complex. This study examines the application of ARIMA-GARCH models in forecasting the USD/JPY exchange rate, focusing on capturing linear trends and time-varying volatility. The study employs data transformations, including logarithmic scaling and differencing, to achieve stationarity and optimize model performance. The ARIMA model identifies and forecasts the conditional mean, while the GARCH model captures volatility clustering in residuals, and rolling window back testing is implemented. Results indicate that the remaining residuals of the ARIMA-GARCH model are white noise and stationary but depart from normality, characteristic of financial time series. The findings demonstrate the utility of ARIMA-GARCH models in modeling the dynamics of foreign exchange rates, including volatility clustering and short-term shocks, while predicting a pattern of depreciation and diminishing volatility, or market stabilization, in the USD/JPY exchange rate. The study offers practical insights for policymakers and market participants managing exchange rate risks in volatile financial environments.

Keywords: ARIMA-GARCH model; exchange rate; time series forecast

1 Introduction

The global foreign exchange market is a liquid financial market. It floats relative to the supply and demand between two or more countries, with influences on international trade, investment, and monetary policy. Exchange rate forecasting is significant in expecting the appreciation or depreciation of interest rates, and helping policymakers, businesses, and investors optimize decisions in regard to federal and independent purchasing power [1]. Exchange rate forecasting is challenging due to the inherent complexity, non-linearity, and

and volatility of such financial time series, embedded in macro and micro social factors. Early studies relied on linear models such as the Autoregressive Integrated Moving Average (ARIMA) model to capture the time-dependent patterns of exchange rates, started by Box and Jenkins in 1970 for time series forecasting, demonstrating how it could be applied to capture linear dependencies [2]. However, financial time series often exhibit volatility clustering, heteroscedasticity, and leptokurtosis (heavy tails), which ARIMA cannot adequately address [3].

To address these deficiencies, models such as the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model were introduced [4]. GARCH models are more focused on time-varying volatility, providing an application to capture periods of heightened or diminished uncertainty in financial markets. In more recent years, hybrid models combining the strengths of ARIMA and GARCH have been explored to overcome the shortcomings of individual approaches. Hybrid ARIMA-GARCH models integrate the linear modeling capabilities of ARIMA for the mean process and the volatility-capturing ability of GARCH for the conditional variance [5, 6]. Recent research has further built upon these foundations; for instance, in the combination of ARIMA models with machine learning models like Long Short-Term Memory (LSTM) networks due to the efficiency of ARIMA for short-term forecasting [7]. ARIMA-GARCH models remain strong in capturing volatility clustering in currency markets, especially when evaluating accuracy through metrics like Root Mean Squared Error (RMSE) [8]. These models have demonstrated better performance in forecasting financial time series compared to standalone models, but empirical applications to exchange rate forecasting remain limited. Furthermore, the development of the models into trading strategies with exchange rates has yet to be accomplished.

This study aims to address these gaps by developing and evaluating an ARIMA-GARCH model to forecast the USD/JPY exchange rate, and furthermore explain potential implications of the model in strategies on the USD/JPY exchange rate.

2 Research Design

This study adopts a combination of quantitative and qualitative research methods. Based on the historical data values of the USD/JPY spot exchange rate, it uses the ARIMA-GARCH model, which explains the mean structure of the model through the ARIMA model, whose residuals are used in the following GARCH model which captures the volatility dynamics.

In the first stage of the modelling process, the ARIMA model is employed to model and forecast the conditional mean of the USD/JPY exchange rate, extracting linear dependencies. In the second stage, the residuals from the ARIMA model are then fitted into a GARCH model to capture volatility dynamics. The combined ARIMA-GARCH model is used to produce forecasts for the mean and variance of the exchange rate, and the

exchange rate forecast is then reconstructed with the results of the model. Lastly, the research assesses the applicability and performance of the ARIMA-GARCH model in forecasting the USD/JPY exchange rate through evaluating model accuracy, with methods such as rolling-window validation against actual exchange rate movements.

2.1 Data and Extraction

In the data collection stage, the historical data of the Japanese Yen to U.S. Dollar Spot Exchange Rate was collected from the Federal Reserve Economic Data, a reliable and authoritative source for information regarding the United States' monetary system, from December 1974 to December 2024 for a period of 50 years and daily frequency.

Firstly, the log of the dataset was taken to stabilize the volatility of the time series, but did not produce a stationary dataset; thus, first-differencing was applied upon the original time series. Upon conducting the Ljung-Box test on the transformed datasets, a combined transformation of both log and first-differencing produced a new dataset with stationarity, with a p-value of 0.0867. The transformed datasets are plotted in Fig. 1:

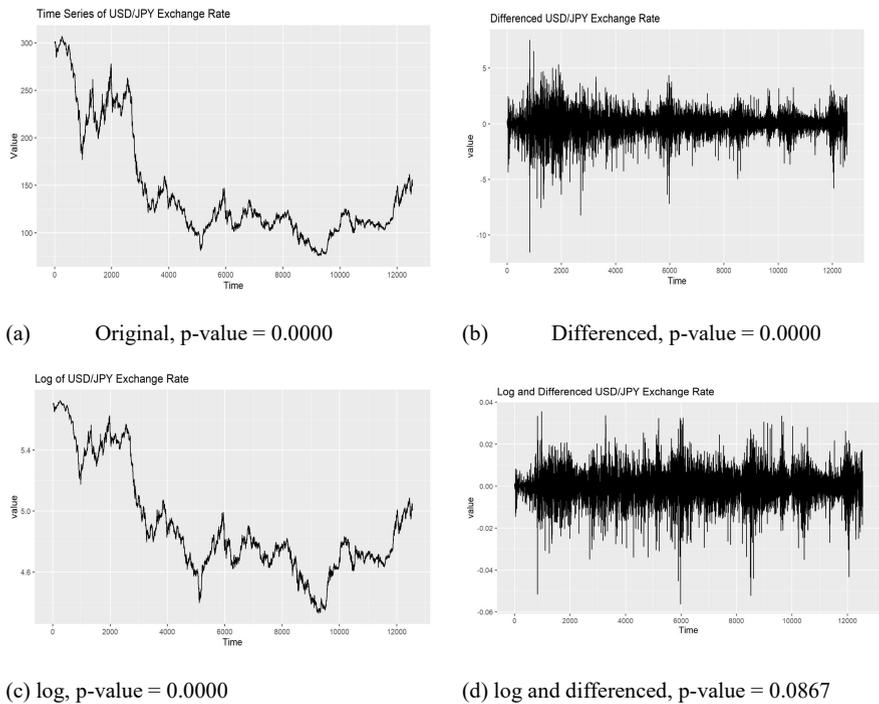


Fig. 1. time series plots and p-values of Ljung-Box tests (Photo/Picture credit: Original).

As observed from Fig. 1 (a), the dataset has no cyclicity or seasonality, except at certain intervals of horizontal trends. The dataset exhibits a general decreasing trend, measured from linear extrapolation between the initial and last data point, but inconsistently; multiple decreases are followed by fluctuational increases. This study will focus mainly on model results from Fig. 1 (c) and Fig. 1 (d), which exhibit volatility clustering where varying levels of volatility are followed by similar levels of volatility and thus ‘cluster.’

2.2 Model Components

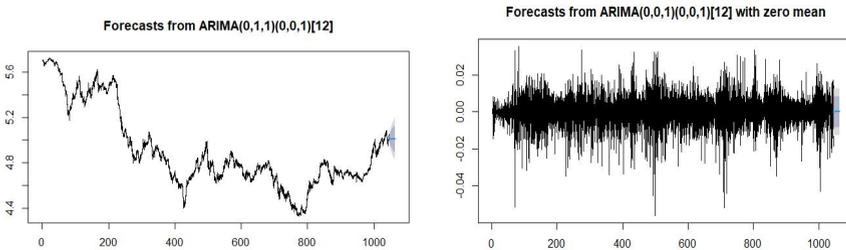
The ARIMA model combines the Autoregressive (AR) model and the Moving Average (MA) model, with additional integrative (I) components. It is typically denoted as ARIMA (p, d, q), where p is the order of the autoregressive (AR) part; d is the degree of differencing, which makes the series stationary; and q is the order of the moving average (MA) part.

The GARCH model, short for Generalized Auto-Regressive Conditional Heteroskedasticity, extends the ARCH model, and incorporates both past error terms and past conditional variances to model time-varying volatility. It is commonly denoted as GARCH (p, q), where p is the order of the lagged conditional variances of the model, and q is the order of the lagged squared residuals in the model; and has three parts, consisting of a mean equation, a variance equation, and the conditional variance.

The ARIMA and GARCH models can be combined to form the ARIMA-GARCH model, where the ARIMA model is used to model the mean process, while the GARCH model is used to model the variance process (volatility).

2.3 Model Fitting

In the model fitting process, the ARIMA models underwent automatic parameter optimization for the dataset. Predictions were generated for 200 days into the future, based on the previous ARIMA model.



- (a) log ARIMA (0, 1, 1)(0, 0, 1)
- (b) differenced log ARIMA (0, 0, 1)(0, 0, 1)

Fig. 2. ARIMA forecasts

Fig. 2 depicts the time series in black plotted against the x-axis with 200-day forecasts -- the extension of the time series with the blue line -- and the value against the y-axis. As shown in Fig. 2, the optimized ARIMA (p, d, q) models are shown to have varying gray confidence intervals, which indicate fluctuating levels of prediction accuracy, but similar linear trends of USD/JPY depreciation.

In the second stage of the modeling process, the GARCH model verifies these forecasts and acts as the variance with the residuals of the ARIMA model. In GARCH specification, the mean model is defined with parameters p and q from the previous ARIMA (p, d, q) model to be ARMA (0, 1) for both transformed time series. To determine parameters p, q of the variance equation, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of squared residuals of the ARIMA model are plotted in Fig. 3:

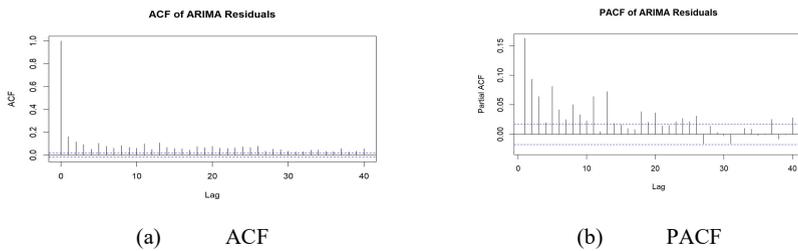


Fig. 3. ACF and PACF of log and differenced log ARIMA Squared Residuals

As observed from Fig. 3 (a), the ACF plot depicts a significant lag at lag 1 that abruptly decreases in magnitude in following lags, thus presenting the GARCH q parameter of 1. Furthermore, from Fig. 3 (b), the PACF depicts a significant lag at lag 1 that decreases substantially, but with continued significant lags that continue until an abrupt magnitude decrease at lag 4, presenting an optimal GARCH p parameter of 3. Thus, the variance equation of the GARCH model is defined to be a standard GARCH (3, 1). Lastly, the distribution follows a normal distribution model, and thus the GARCH models for both transformed time series are essentially specified with the same parameters.

Next, the GARCH plots are generated, of which the returns series are plotted against 1% Value at Risk (VaR) limits:

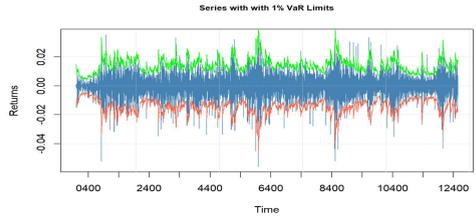
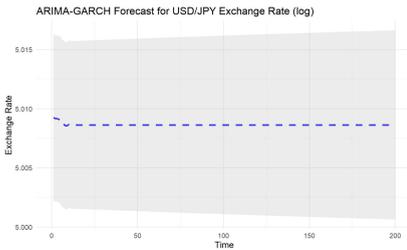


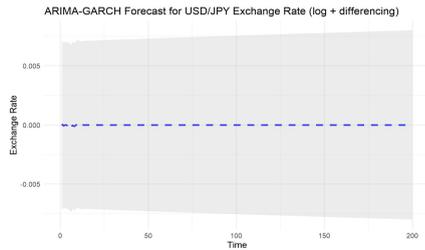
Fig. 4. Returns series against 1% VaR limits

In Fig. 4, the blue represents the returns of the dataset plotted against the time on the x-axis, and the green and red boundaries represent gains or losses respectively. The 1% VaR boundaries are based on the estimated conditional standard deviation, the estimated volatility of a time series at a specific point in time, conditional on past information. The return volatility of both time series are equivalent, thus resulting in an identical 1% VaR plot. Fig. 4 superimposes the estimated volatility on the observed returns to highlight how the model explains the dynamics of market fluctuations; wider bands indicate periods of high volatility, while narrower bands indicate periods of stability. As seen from Fig. 4, these events occur frequently due to fluctuating limits. Fig. 4 furthermore illustrates the volatility clustering within the time series, as observed from volatility persistence; furthermore, VaR limits are observed to take the shape of waves, where consistent volatility increases are followed by consistent volatility decreases, indicating mean-reversion.

What is also generated are the unconditional sigma predictions fitted onto the forecasting GARCH model. The sigma predictions forecast the unconditional variability in level of the time series, which represents the average, or long-term volatility, without rolling forecasts. These GARCH volatility predictions are applied upon the original ARIMA models to act as forecast boundaries:



(c) dataset with log



(d) dataset with log and differencing

Fig. 5. Combined mean and variance forecasts

In Fig. 5, the 200-day forecast window is plotted against the x-axis, and the ARIMA forecast in blue dashed lines plotted against the y-axis with gray boundaries calculated based on the ARIMA forecast plus or minus the predicted GARCH volatility. As observed from Fig. 5, initial fluctuations indicative of the influence of recent shocks or volatility clusters in the historical data stabilize over time.

In model evaluation, rolling-window backtesting was used. The model repeatedly performs fitting on decreasing amounts of training data, and multiple comparisons are performed between actual historical data for the window period and forecasting validation data. The data from the train window was fitted onto the previously specified ARIMA models, and forecasts were generated every $\frac{1}{5}$ of a year on a validation window of 50 days, with the time frame shifted forward to repeat 200 times starting from a 10-year benchmark. This procedure was performed on Fig. 1 (c) to act as a metric to assess predictive accuracy over time, as it is commonly used in non-stationary financial data to create predictions iteratively.

3 Results

The results from the ARIMA and ARIMA-GARCH modeling process provide an assessment of the USD/JPY exchange rate's future behavior. The ARIMA-GARCH models were evaluated based on their parameter combinations and performance metrics, shown in Table 1:

Table 1. Accuracy measures for ARIMA-GARCH models for each of four datasets

Index	RMSE (whole)	RMSE (train)	RMSE (test)
ARIMA (0, 1, 1)	0.0064	0.3584	0.0893
ARIMA (0, 0, 1)	0.0065	0.0064	0.0071

The ARIMA-GARCH models were effective in capturing general trends in the USD/JPY exchange rate, with forecasts consistently predicting a pattern of depreciation and increasing stability evidenced by reduced volatility in predictions. From Table I, it can be observed that ARIMA (0, 0, 1) had a lower RMSE value compared to ARIMA (0, 1, 1), indicative of better performance for shorter-term predictions.

Furthermore, the incorporation of GARCH models modeled the variance (volatility) of the residuals, highlighting that unconditional sigma (long-term volatility) stabilizes over time, consistent across both data transformations. The initial fluctuations in predicted volatility are indicative of recent historical shocks, which diminish as the forecasts extend into the future, supporting the expectation of long-term stability.

Upon evaluation of the final ARIMA-GARCH model on both transformed datasets, multiple tests were performed upon the residuals of the model to detect white noise, normality, and stationarity:

Table 2. P-values of Ljung-Box, Kolmogorov-Smirnov (K-S), and Augmented Dickey-Fuller (ADF) tests

Index	Ljung-Box	Kolmogorov-Smirnov	Augmented Dickey-Fuller
GARCH (3, 1)	0.2267	0.0100	0.0000

The values from Table 2 produced the following observations, proving the usefulness of the GARCH model in capturing remaining variance such as time-varying volatility clustering in comparison to individual ARIMA models:

1. The residuals are confirmed to be a white noise series with a p-value greater than 0.05.
2. The residuals are stationary, with a p-value less than 0.05, suggesting that the model captures the time series dynamics well.
3. The residuals do not follow a normal distribution, with a small p-value less than 0.05. However, this is common in datasets that are large or follow heavy-tailed, rather than normal distributions, such as financial time series.

Lastly, the results of the rolling-window backtesting were plotted, and calculations performed cumulating the total number of occurrences where historical values exceeded 95% confidence intervals:

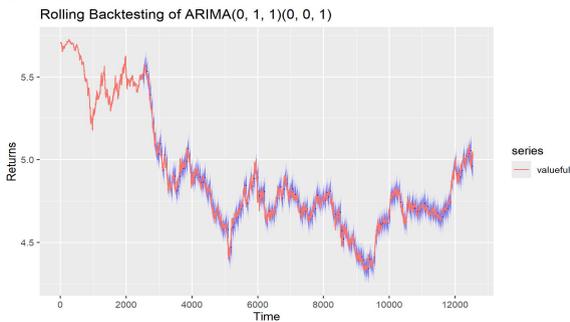


Fig. 6. 200 rolling backtesting of log ARIMA-GARCH model with 50-day forecasts

Fig. 6 plots the results of the rolling backtesting for the log dataset, with the red line representing historical value and blue areas representing confidence intervals of each 50-day forecast. From Fig. 6, it can be observed that prediction intervals remain relatively consistent except in areas with increased volatility that indicate higher uncertainty; overall, the model is able to capture the time series' general directionality. Out of the 200

backtesting forecasts, 57 cases were identified as true positives where historical data exceeded 95% confidence intervals, and 143 cases as negatives, supporting results with a probability of 71.5% in which the model can accurately model the range of future price.

4 Discussion

The results align with previous research in the utilization of ARIMA-GARCH models on forecasting exchange rates. GARCH models have advanced ability to model time-varying volatility in financial time series [4, 9]. Furthermore, the finding that the volatility forecasts converge to a stable long-term level supports theoretical expectations of decreasing uncertainty as markets stabilize over time [10].

The study indicates how the ARIMA-GARCH model can be implemented in real-world market decision-making. The forecasted volatility from GARCH can be used to calculate risk management and appropriate stop-loss or take-profit levels for the USD/JPY exchange rate. It may be used to forecast future price ranges, and strategies may be developed based on forecast outcomes, for example, a stop-loss strategy whenever price falls below the forecast interval, due to the assumption of the 73% probability derived from rolling-window backtesting. The model can be continuously adapted to the latest data, in a process similar to rolling-window backtesting, ensuring forecast relevance as market conditions evolve.

The ARIMA-GARCH model, however, has limitations. First, the model assumes stationary residuals and normality in GARCH innovations, which may not fully capture the heavy tails and non-linear dependencies present in real-world exchange rates [11]. Secondly, while the ARIMA-GARCH framework is effective for short-term forecasting, it may have limited applicability to long-term forecasting, as it does not incorporate exogenous variables like trade balances or economic indicators.

Future research should continue to explore hybrid models alongside traditional ARIMA-GARCH methods to capture non-linear patterns and improve long-term forecasts. For example, GARCH-EGB2 models can be experimented in relation to ARIMA models, as they capture many traditional characteristics of exchange rates [3]. Furthermore, ARIMA-GARCH models can be closer examined where prices go out of forecasted range. Rolling-window backtesting can provide helpful insights on if the model is systematically underestimating in certain conditions, or if any patterns exist in out-of-range occurrences such as sudden turns in trends.

5 Conclusion

This study analyzed the USD/JPY exchange rate using ARIMA and GARCH models to capture linearity and volatility, respectively. The analysis demonstrated that combining

these models effectively models the exchange rate's dynamics, such as volatility clustering and short-term shocks, while providing predictions of long-term stability. GARCH forecasts revealed a pattern of volatility clustering that diminished over time, indicating market stabilization, and model residuals were re-tested that determined the success of the model in capturing the variance ARIMA had failed to capture. These findings contribute to the existing knowledge by exhibiting the effectiveness of ARIMA-GARCH models in forecasting volatile time series and modeling exchange rate dynamics.

The implications of the findings suggest that policymakers and market participants can use ARIMA-GARCH models to better understand exchange rate movements and associated risks. For example, the predicted stability in volatility suggests reduced future risks in the USD/JPY market, providing a basis for informed decision-making. However, although this study uses historical data to build an effective model, the models rely on assumptions of stationarity and normality, and their effectiveness is limited to short-term forecasting and lacks incorporation of exogenous economic factors. Although historical data can provide a certain reference, the market is dynamic and future prices and volatility may be affected by unpredictable factors.

These insights have practical applications for policymakers and market participants in managing exchange rate risks. Expanding its integration with machine learning techniques or economic indicators could further improve its utility in dynamic financial environments. With further advancements, ARIMA-GARCH models could be integrated into automated trading systems or risk management frameworks.

References

1. Yoshikawa, H.: On the Equilibrium Yen-Dollar Rate. *American Economic Review* 80(3), 576–583 (1990).
2. Box, G. E. P., Jenkins, G. M.: *Time series analysis: Forecasting and control*. Rev. edn. Holden-Day, San Francisco (1976).
3. Wang, K., Fawson, C., Barrett, C. B., McDonald, J. B.: A flexible parametric GARCH model with an application to exchange rates. *Journal of Applied Econometrics* 16(4), 521–536 (2001).
4. Bollerslev, T.: Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31(3), 307–327 (1986).
5. Yu, J., Wang, H., Lai, K. K.: Forecasting Exchange Rates Using a Combined ARIMA and GARCH Model. *Applied Financial Economics* 15(1), 19–28 (2005).
6. Wang, J., Wang, F., Zhang, J.: Stock market forecasting using a hybrid model. *Applied Soft Computing* 36, 1–12 (2016).
7. García, F., Guijarro, F., Oliver, J., Tamošiūnienė, R.: Foreign Exchange Forecasting Models: ARIMA and LSTM Comparison. In: *9th International Conference on Time Series and Forecasting* (2023).
8. Adenomon, M. O., Emmanuel, J. R.: Application of ARIMA-GARCH Models in Volatility Forecasting. *JRSS-NIG* (2024).

9. Engle, R. F.: Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation. *Econometrica* 50(4), 987–1007 (1982).
10. Mussa, M.: Nominal exchange rate regimes and the behavior of real exchange rates: Evidence and implications. *Carnegie-Rochester Conference Series on Public Policy* 25(1), 117–214 (1986).
11. Cont, R.: Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1(2), 223–236 (2001).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

