Exchange Rate Analysis Between the U.S. Dollar and the Japanese Yen

Yuta Sakiyama
Exchange Rate Analysis Between the U.S. Dollar and the Japanese Yen

Yuta Sakiyama
September, 2023

Contents

1 Introduction 2

2 Model and Theory 3
   2.1 Stochastic Process and Stationary Process . . . . . . . . . . . . . 3
   2.2 ACF and PACF . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
   2.3 AIC and BIC . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
   2.4 MAE, RMSE and Ljung-Box Test . . . . . . . . . . . . . . . . . . 8
   2.5 AR Model, MA Model, ARMA Model and ARIMA Model . . . . . 10
   2.6 ARCH Model, GARCH Model and Fractional Differencing Model 12

3 Result 13
   3.1 Investigations of Data Properties . . . . . . . . . . . . . . . . . . 13
   3.2 Result of ARIMA Model . . . . . . . . . . . . . . . . . . . . . . . 14
   3.3 Result of GARCH Model and Fractional Differencing Model . . . 14

4 Conclusion 27

5 Acknowledgements 27

6 Appendix 27
   6.1 Python Programming Code . . . . . . . . . . . . . . . . . . . . . 27
   6.2 R Programming Code . . . . . . . . . . . . . . . . . . . . . . . . 28

7 References 33

Abstract
The exchange data between the U.S. Dollar and Japanese Yen are analyzed with three models called the Auto-Regressive Integrated Moving-Average (ARIMA) model, the Generalized Auto-Regressive Conditional Heteroscedastic (GARCH) model, and the Fractional Differencing model. We mainly use log-transformed data with one difference taken because the residuals of the ARIMA model with log-transformed data is closer to
normal distribution than other residuals of ARIMA models with square-root-transformation or without transformation. Furthermore, the Akaike Information Criteria (AIC), the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) are used to evaluate performances of each models. The Bayesian Information Criterion (BIC) is also used for GARCH models and we do the Ljung-Box tests for ARIMA models. In conclusion, the fractional differencing model is the best one in the three types of models because its MAE and its RMSE are the smallest values in those of all models we make in this analysis.

1 Introduction

In financial studies, it seems mostly inevitable to analyze time series data because getting maximum wealth is the biggest goal to study finance with data which has changed over time. So it is important for its study to research how to predict the future situation by analyzing time series data. For this reason, we focus on time series analysis as one statistical method to study one part of finance in this paper.

However, there are lots of data related to finance like stock prices, exchange rates, trends in insurance and so on. From many fields to study, we choose the Japanese Yen (JPY) exchange rate against the U.S. Dollar (USD) to study the time series analysis with the financial application. There are two reasons to select this data.

The first reason is that the exchange rate between the U.S. Dollar and the Japanese Yen seems to have been fluctuating wildly against the Dollar recently. On 20th October in 2022, the Japanese Yen hit a record low against the U.S. Dollar marking the first time in 32 years since August 1990. It means that 1 dollar was equal to 150 yen, but we could buy 1 dollar for about 115 yen at the beginning of 2023. In this situation, it is questionable whether it was possible to predict the future rate with statistical methods before the soaring U.S. Dollar and what the future rate will be by using time series analysis. (Sumitomo Mitsui DS Asset Management Company, 2022)

The second reason is that the yen has been under the floating exchange rate system against the dollar. In other words, it also means that the fluctuation range is not intentionally restricted by government policy. If there are some arbitrary factors to change the rate, we have to add some economical ideas to this analysis. But this paper is a bachelor thesis on mathematics and it is better to focus on mathematical discussion rather than an economical one. So, it makes sense to choose this topic to write the mathematical thesis with the economical application.

To analyze this data, we will use three types of models; the Auto-Regressive Integrated Moving-Average (ARIMA) model, the Generalized Auto-Regressive Conditional Heteroscedastic (GARCH) model, and the Fractional Differencing model. The ARIMA model is made by the Auto-Regressive Moving-Average (ARMA) model, which is one of the most important linear processes. The GARCH model and the Fractional Differencing model are non-linear processes.
Compared to the ARIMA model, the former model can adapt to changes in volatility in finance with the assumption of variable conditional variance, and the latter one is a long memory time series which can avoid too severe a modification into predicted values. We analyze the exchange data with those models then see what predictions they make.

2 Model and Theory

2.1 Stochastic Process and Stationary Process

First, we present definitions of a probability space and a stochastic process before stating definitions of two types of stationary processes.

**Definition 1** (Probability Space). Let us define the sample space as the set of all possible outcomes of an experiment which is written as \( \Omega \). And we also define a collection \( \mathcal{F} \) of subsets of \( \Omega \) as a \( \sigma \)-field if it satisfies the following three conditions:

(a) \( \emptyset \in \mathcal{F} \);

(b) if \( A_1, A_2, \ldots \in \mathcal{F} \) then \( \bigcup_{i=1}^{\infty} A_i \in \mathcal{F} \);

(c) if \( A \in \mathcal{F} \) then \( A^c \in \mathcal{F} \).

Additionally, a probability measure \( P \) on the pair \( (\Omega, \mathcal{F}) \) is a function \( P : \mathcal{F} \to [0, 1] \) satisfying

(a) \( P(\emptyset) = 0, \; P(\Omega) = 1; \)

(b) if \( A_1, A_2, \ldots \) is a collection of disjoint members of the collection \( \mathcal{F} \), that is \( A_i \cap A_j = \emptyset \) for all pairs \( i, j \) satisfying \( i \neq j \), we say

\[
P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).
\]

The triple \( (\Omega, \mathcal{F}, P) \) is called a probability space. (Grimmett and Stirzaker, 2020)

**Definition 2** (Stochastic Process). A stochastic process on the probability space \( (\Omega, \mathcal{F}, P) \) is a family of random variables \( Y(t, \omega) \), which can be written as:

\[
\{ Y(t, \omega) : t \in \mathcal{T}, \omega \in \Omega \}
\]

where \( \mathcal{T} \) is a set of points of time. \( Y(t, \cdot) \) with \( t \) fixed is a random variable and \( Y(\cdot, \omega) \) with \( \omega \) fixed is called a sample path. We abbreviate \( Y(t, \omega) \) to \( \{y_t\} \) without \( \omega \) written and we specify realizations from the family \( Y(t, \omega) \) in that abbreviation. (Tanaka, 2020)
In time series analysis, when $\mathcal{T}$ is the set of integers $\mathbb{Z}$, the observed time series data \( \{y_t\}_{t=1}^n \) is thought as realizations from a sequence of random variables \( \{y_t\}_{t \in \mathbb{Z}} \), then we assume a property or a structure related to generating processes of the time series \( \{y_t\}_{t=1}^n \). A structure of a stochastic process is called a time series model, so the stochastic process can be meant to be the time series data with some characteristics. Hence let the time series \( \{y_t\} \) denote the stochastic process hereafter. (Shimada, 2019)

The next two definitions show us the most fundamental characteristics for the stochastic process and to analyze time series data. We fix the set $\mathcal{T}$ to the set of integers $\mathbb{Z}$ because the exchange data used in this study is taken per one day except for Japanese weekends.

**Definition 3 (Strictly Stationary Process).** Let \( \{y_t\} \) be a stochastic process. For any $\forall n \in \mathbb{N}$, for any $\forall (y_1, ..., y_n) \in \mathbb{R}^n$, and for any $\forall t_1, ..., \forall t_n, \forall h \in \mathbb{Z}$, the stochastic process \( \{y_t : t \in \mathbb{Z}\} \) is a strictly stationary process if the below equation holds:

$$F_{t_1, t_2, ..., t_n}(y_1, y_2, ..., y_n) = F_{t_1+h, t_2+h, ..., t_n+h}(y_1, y_2, ..., y_n)$$

(2)

where

$$F_{t_1, t_2, ..., t_n}(y_1, y_2, ..., y_n) := P(Y(t_1, \cdot) \leq y_1, ..., Y(t_n, \cdot) \leq y_n)$$

(3)

is the joint distribution function (Shiraishi, 2022).

This definition means that all joint distribution functions depend only on the lag $h$ under the strictly stationary process. However, all strict stationary processes are not weakly stationary.

**Definition 4 (Weakly Stationary Process).** Let \( \{y_t\} \) be a stochastic process again. For any $t, z \in \mathbb{Z}$, if the mean $E[y_t]$ and the covariance $\text{Cov}[y_t, y_{t+h}]$ exist and the following two conditions are satisfied, the stochastic process \( \{y_t\} \) is a weakly stationary process:

$$E[y_t] = \mu$$

(4)

$$\text{Cov}[y_t, y_{t+h}] = R(h).$$

(5)

Note that $\mu$ doesn’t depend on $t$ and $R(h)$ is a function which depends only on the lag $h$. The function $R(\cdot)$ is defined as the autocovariance function (Tanaka, 2020).

In some books like Brockwell and Davis, 2006, and Shiraishi, 2022, there is an additional condition as $E[|X_t|^2] < \infty$. In this thesis we do not think about this condition but it is needed to discuss the time series on the Hilbert spaces.
2.2 ACF and PACF

We present tools to analyze the strength of relationships between time series data in this section.

**Definition 5** (Auto-Correlation Function, ACF). Let \( R(h) = \text{Cov}[y_t, y_{t+h}] \) be the autocovariance function of the stationary process. Here we can define the correlation coefficient of \( y_t \) and \( y_{t+h} \) as

\[
\rho(h) = \frac{\text{Cov}[y_t, y_{t+h}]}{\sqrt{\text{Var}[y_t]} \sqrt{\text{Var}[y_{t+h}]}}. \tag{6}
\]

This is called an Auto-Correlation Function, an ACF. As in the above definition, this is a function only depending on the lag \( h \). Additionally, we can describe the autocorrelation function of a stationary process like

\[
\rho(h) = \frac{R(h)}{R(0)}. \tag{7}
\]

The ACF \( \rho(h) \) shows us the relationship strength between \( y_t \) and \( y_{t+h} \). But mostly \( y_{t+h} \) is related to not only \( y_t \) but also \( y_{t+h-1}, \ldots, y_{t+1} \), so the ACF \( \rho(h) \) includes also influences from \( y_{t+h-1}, \ldots, y_{t+1} \). To remove the indirect influences from \( y_{t+h-1}, \ldots, y_{t+1} \), we define another function called the PACF, which is discussed below. (Shiraishi, 2022)

**Definition 6** (Partial Auto-Correlation Function, PACF). Let us assume that \( y_t \) is a stationary process with \( \mu \) as the average. Additionally, for \( h \geq 2 \), let \( \hat{y}_{t+h} \) denote the regression of \( y_{t+h} \) on \( \{y_{t+h-1}, y_{t+h-2}, \ldots, y_{t+1}\} \). This is written as follows:

\[
\hat{y}_{t+h} = \beta_1 y_{t+h-1} + \beta_2 y_{t+h-2} + \ldots + \beta_{h-1} y_{t+1}. \tag{8}
\]

For \( 1 \leq j \leq t+1 \), \( \beta_j \) is defined to satisfy the condition that \( \hat{y}_{t+h} \) minimizes the mean squared error \( E[(y_{t+h} - \sum_{j=1}^{h-1} \beta_j y_{t+h-j})^2] \). Furthermore, let \( \hat{y}_t \) denote the regression of \( y_t \) on \( \{y_{t+1}, y_{t+2}, \ldots, y_{t+h-1}\} \), which we write as below:

\[
\hat{y}_t = \beta_1 y_{t+1} + \beta_2 y_{t+2} + \ldots + \beta_{h-1} y_{t+h-1}. \tag{9}
\]

This equation is defined in the same way for \( \hat{y}_{t+h} \). The coefficients, \( \beta_1, \ldots, \beta_{h-1} \) are also the same in both equations (8) and (9) because of stationarity. Then, we can describe the following equation with the unknown real number sequence \( \phi_{h1}, \phi_{h2}, \ldots, \phi_{hh} \):

\[
\phi_{h1} = \text{Corr}(y_{t+1}, y_t) = \rho(1), \quad h = 1 \tag{10}
\]

and

\[
\phi_{hh} = \text{Corr}(y_{t+h} - \hat{y}_{t+h}, y_t - \hat{y}_t), \quad h \geq 2 \tag{11}
\]
where $\text{Corr}(\cdot, \cdot)$ means the correlation calculation as:

$$
\phi_{hh} = \frac{\text{Cov}[y_{t+h} - \hat{y}_{t+h}, y_t - \hat{y}_t]}{\sqrt{\text{Var}[y_{t+h} - \hat{y}_{t+h}]} \sqrt{\text{Var}[y_t - \hat{y}_t]}}, \quad h \geq 2.
$$

The value of $\phi_{hh}$ in the equations is called the Partial Auto-Correlation of $y_t$ and $y_{t+h}$. (Shumway and Stoffer, 2017)

If the process $y_t$ is Gaussian, we can write $\phi_{hh}$ as

$$
\phi_{hh} = \text{Corr}(y_{t+h}, y_t | y_{t+1}, \ldots, y_{t+h-1}).
$$

Furthermore, Gaussian processes means that each time series $y_1, y_2, \ldots, y_n$ has the multivariate normal distribution. As in the statement of the ACF, the PACF shows us the strength of the linear relationship between $y_t$ and $y_{t+h}$ by the removal of the influences from $y_{t+h-1}, \ldots, y_{t+1}$. (Shumway and Stoffer, 2017)

The PACF can actually be expressed as the solution of the below formula:

$$
\begin{pmatrix}
1 & R(1) & \ldots & R(h-1) \\
R(1) & 1 & \ldots & R(h-2) \\
\vdots & \vdots & \ddots & \vdots \\
R(h-1) & R(h-2) & \ldots & 1 \\
\end{pmatrix}
\begin{pmatrix}
\phi_{h1} \\
\phi_{h2} \\
\vdots \\
\phi_{hh} \\
\end{pmatrix}
=
\begin{pmatrix}
R(1) \\
R(2) \\
\vdots \\
R(h) \\
\end{pmatrix}.
$$

(Tanaka, 2020)

Finally, we define the function called the sample ACF. This is not used for analyzing the exchange data directly, but used to discuss the Ljung-Box test later.

**Definition 7** (Sample Auto-Correlation Function, Sample ACF). The sample autocorrelation function $\hat{\rho}(h)$ is defined as below:

$$
\hat{\rho}(h) = \frac{\hat{R}(h)}{\hat{R}(0)}
$$

where

$$
\hat{R}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (y_{t+h} - \mu)(y_t - \mu), \quad \mu = \frac{1}{n} \sum_{t=1}^{n} y_t.
$$

**2.3 AIC and BIC**

We begin the discussion of linear regression to explain two information criteria, the AIC and the BIC. We express the linear regression model as below:

$$
y_t = \beta_0 + \beta_1 z_{t1} + \beta_2 z_{t2} + \ldots + \beta_q z_{tq} + w_t.
$$
In the equation (17), $\beta_0, \beta_1, ..., \beta_q$ are unknown fixed regression coefficients and $\{w_t\}$ is a random error or the noise process consisting of independent and identically distributed normal valiable with mean 0 and variance $\sigma^2_w$. When we define the column vectors $z_t = (1, z_{t1}, z_{t2}, ..., z_{tq})'$ and $\beta = (\beta_0, \beta_1, ..., \beta_q)'$ which $(\cdot, ..., \cdot)'$ from the vector $(\cdot, ..., \cdot)$ denotes transpose, we can rewrite the equation (17) as follows:

$$y_t = \beta_0 + \beta_1 z_{t1} + \beta_2 z_{t2} + ... + \beta_q z_{tq} + w_t = \beta' z_t + w_t $$

(18)

where $w_t \overset{i.i.d.}{\sim} N(0, \sigma^2_w)$ and i.i.d. means to be independent and identically distributed. By the equation (18), we can write the error sum of squares

$$Q = \sum_{t=1}^{n} w_t^2 = \sum_{t=1}^{n} (y_t - \beta' z_t)^2. $$

(19)

With the ordinary least square estimation, denoted the OLS estimation, we obtain the vector $\hat{\beta}$ that minimizes the error sum of square $Q$ with respect to $\beta_0, \beta_1, ..., \beta_q$, that is the vector $\beta$. Now we try to find the concrete form of the vector $\hat{\beta}$. This minimization with this estimation can be accomplished by differentiating the equation (19) with respect to the vector $\beta$ to find the local minimum of the parabola or by using the properties of projections. Either way, we get the solution as

$$\sum_{t=1}^{n} (y_t - \hat{\beta}' z_t) z_t' = 0. $$

(20)

Additionally, with a transformation, we obtain

$$\sum_{t=1}^{n} \hat{\beta}' z_t z_t' = \sum_{t=1}^{n} y_t z_t'. $$

(21)

Then transposing the vectors of both sides gives us the normal equations

$$\left(\sum_{t=1}^{n} z_t z_t'\right) \hat{\beta} = \sum_{t=1}^{n} z_t y_t. $$

(22)

If $\sum_{t=1}^{n} z_t z_t'$ is non-singular, we can also write like

$$\hat{\beta} = \left(\sum_{t=1}^{n} z_t z_t'\right)^{-1} \sum_{t=1}^{n} z_t y_t. $$

(23)

Now we understand the shape of the vector $\hat{\beta}'$ specifically. Back to the story, we change $\beta$ to $\hat{\beta}$ in the equation(19), then we obtain the minimized error sum of squares, denoted SSE, as follows:

$$SSE = \sum_{t=1}^{n} (x_t - \hat{\beta}' z_t)^2. $$

(24)
Finally, we define the maximum likelihood estimator as

$$\hat{\sigma}_q^2 = \frac{SSE(q)}{n}$$  \hspace{1cm} (25)

where $q$ is the number of parameters in the model except for the constant term $\beta_0$, that is the number of coefficients in the equation (17). The SSE is regarded as the function with respect to $q$.

As an example, by the way, the above discussion is progressed by using linear regression. But it can be relaxed for the time series models which we define later by changing the variables of the equation (17). (Shumway and Stoffer, 2017)

Now we get all preparations to define the AIC and the BIC.

**Definition 8** (Akaike Information Criterion, AIC). We define the Akaike Information Criterion, the AIC as

$$AIC = \log \hat{\sigma}_q^2 + \frac{n + 2q}{n}.$$  \hspace{1cm} (26)

The value of $q$ yielding the minimum AIC indicates the best model. We must use this measure with models that build on the same transformation. (Shumway and Stoffer, 2017)

**Definition 9** (Bayesian Information Criterion, BIC). We also define the Bayesian Information Criterion, the BIC as

$$BIC = \log \hat{\sigma}_q^2 + \frac{q \log n}{n}.$$  \hspace{1cm} (27)

The BIC is also called the Schwarz Information Criterion, the SIC. It is also not valid to use this measure with models that are made on different transformations. (Shumway and Stoffer, 2017)

### 2.4 MAE, RMSE and Ljung-Box Test

We understand the AIC and the BIC by the above discussion. However, they cannot be used to evaluate all models in this thesis actually. So it is better to have any means to be applied for all of them. The simple ones are to measure the distances between predicted values and actual values then take the sum of them. The below two are based on this idea.

**Definition 10** (Mean Absolute Error, MAE). The mean absolute error is defined as

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |\hat{x}_t - x_t|$$  \hspace{1cm} (28)

where $\hat{x}_t$ is a predicted value at time $t$ and $x_t$ is an actual value at time $t$. (Nura, Doguwa, and Yusuf Basiru, 2020)
**Definition 11** (Root Mean Square Error, RMSE). The root mean square error is defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (\hat{x}_t - x_t)^2}$$

(29)

where $\hat{x}_t$ and $x_t$ are the same as the MAE. (Nura, Doguwa, and Yusuf Basiru, 2020)

We will use these measures to evaluate model performances. We can think that the less value, the better model.

To verify the performance of ARIMA models, we define the Ljung-Box test as follows.

**Definition 12** (Ljung-Box Test). For any fixed number $m \in \mathbb{N}$, we test

Null hypothesis $H_0: \rho(1) = ... = \rho(m) = 0$  

\vspace{1mm}

v.s.

Alternative hypothesis $H_1: \exists k \in \{1, ..., m\}$ \hspace{1mm} s.t. $\rho(k) \neq 0$.  

\vspace{1mm}

The test statistic is given by

$$\tilde{Q} := n(n+2) \sum_{k=1}^{m} \frac{\hat{\rho}(k)^2}{n-k}$$

(32)

where $\hat{\rho}(k)$ is the sample ACF of the residuals $\hat{U}_1, ..., \hat{U}_n$. As $\alpha$ is the significance level, we judge

$$\tilde{Q} > \chi^2_{m-p}(\alpha) \implies \text{rejection}$$

(33)

$$\tilde{Q} \leq \chi^2_{m-p}(\alpha) \implies \text{acceptance}$$

(34)

where $\chi^2_{m-p}(\alpha)$ is the upper $\alpha$ point of the chi-square distribution with the degree of freedom $m - p$, that is $P(X \leq \chi^2_{m-p}(\alpha)) = 1 - \alpha$ if $X$ follows the chi-square distribution $\chi^2_{m-p}$. (Shiraishi, 2022)

In the above definition, $n$ is the sample size and the value of $m$ is chosen somewhat arbitrarily, so we set $m = 20$. The "tsdiag" function in the programming actually shows us the $p$-value of this statistic visually and therefore we can judge the test through checking points above the blue line.
2.5 AR Model, MA Model, ARMA Model and ARIMA Model

In this section, we first define white noise for discussing the ARIMA model.

**Definition 13 (White Noise).** For any $\forall t \in \mathbb{Z}$, if the time series $\{y_t\}$ satisfies the following condition, then the series $\{y_t\}$ is white noise satisfying

$$E[y_t] = 0, \quad \text{Cov}[y_t, y_{t+h}] = \left\{ \begin{array}{ll} \sigma^2 & (h = 0) \\ 0 & (h \neq 0) \end{array} \right. \quad (35)$$

where $E[y_t]$ is the mean of $\{y_t\}$ and $\text{Cov}[y_t, y_{t+h}]$ is the covariance between $\{y_t\}$ and $\{y_{t+h}\}$.

It is clear that white noise is weakly stationary process because both the mean and the variance are constants and the covariance does not depend on $t$.

We next define the AR model definition by using a white noise, denoted $\{\epsilon_t\}$.

**Definition 14 (AR($p$) Model).** Let $\epsilon_1, \ldots, \epsilon_n$ be white noise with mean 0 and variance $\sigma^2$. The Auto-Regressive (AR) model of $p$-order is defined as follows:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t, \quad t = 1, \ldots, n. \quad (36)$$

We shorten $p$-order Auto-Regressive model to say AR($p$) model. (Nomura, 2016)

As this formula shows, the observation value $y_t$ is determined by the past observation values $y_{t-1}, \ldots, y_{t-p}$. The coefficients $\phi_1, \phi_2, \ldots, \phi_p$ are called AR coefficients. In this paper, we define the AR($p$) model without constant term $\phi_0$ although there are some cases to put it on the equation, because we can redefine the time series by subtracting the mean from the equation and cancel $\phi_0$ on the model.

The AR($p$) model is not always weakly stationary. A necessary and sufficient condition that it is weakly stationary, is that the absolute values of the solutions of the characteristic equation

$$\phi_1 B + \phi_2 B^2 + \ldots + \phi_p B^p = 1$$

are smaller than 1. The necessary and sufficient condition that the AR(1) model is weakly stationary, for example, is $-1 < \phi_1 < 1$. If $\phi_1 = 1$, the AR(1) model becomes a random walk. The random walk is a process for which its covariance does not converge a constant where $t$ becomes large. (Nomura, 2016)

**Definition 15 (MA($q$) Model).** Let $\epsilon_1, \ldots, \epsilon_n$ be white noise with mean 0 and variance $\sigma^2$ again. The Moving-Average (MA) model of $q$-order is defined as follows:
\[ y_t = \epsilon_t + \lambda_1 \epsilon_{t-1} + \ldots + \lambda_q \epsilon_{t-q}, \quad t = 1, \ldots, n. \]  

(38)

We also shorten $q$-order Moving-Average model to say MA($q$). (Nomura, 2016)

As this formula shows, the observation value $y_t$ is determined by the past white noises including one at time $t$ like $\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-q}$. The coefficients $\lambda_1, \ldots, \lambda_q$ are called MA coefficients. There are also some cases to put a constant term $\lambda_0$ on this equation like the AR($p$) model, but we will use this model without constant term $\lambda_0$ because of the same reason as AR($p$) model.

The MA($q$) model is always weakly stationary because the sum of white noises gives us what both mean and covariance of that model are constants. (Nomura, 2016)

**Definition 16** (ARMA($p, q$) Model). We saw the AR($p$) model and the MA($q$) model so far. If they are synthesized like the equation below, we can get a new model called the ($p, q$) order Auto-Regressive Moving-Average model. This can be abbreviated and called the ARMA($p, q$) model.

\[ y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t + \lambda_1 \epsilon_{t-1} + \ldots + \lambda_q \epsilon_{t-q} \]  

(39)

The coefficients, $\phi_1, \ldots, \phi_p$ and $\lambda_1, \ldots, \lambda_q$ are also called the AR coefficients and the MA coefficients, respectively. Moreover, we don’t put the constant on the right side. (Nomura, 2016)

**Definition 17** (ARIMA($p, d, q$) Model). To analyze the non-stationary time series by using the ARMA model, we can take one difference of them like $\Delta^1 y_t = y_t - y_{t-1}$ where $d = 1$. In general, after taking $d$ differences $\Delta^d x_t$ for time series $x_t$ until getting an approximate stationary time series, we can put that time series into the ARMA($p, q$) framework. It is called the Auto-Regressive Integrated Moving-Average model, the ARIMA model for short. Its formula is written as

\[ \Delta^d y_t = \sum_{i=1}^{p} \phi_i \Delta^d y_{t-i} + \Delta^d \epsilon_t + \sum_{j=1}^{q} \lambda_j \Delta^d \epsilon_{t-j} \]  

(40)

where

\[ \Delta^1 y_t = y_t - y_{t-1} \]
\[ \Delta^2 y_t = \Delta^1 y_t - \Delta^1 y_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \]
\[ \vdots \]
\[ \Delta^d y_t = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1}. \]

Notationally, it is written as the ARIMA($p, d, q$) model (Shimada, 2019).
2.6 ARCH Model, GARCH Model and Fractional Differencing Model

**Definition 18 (ARCH Model).** When the time series \( \{y_t\} \) satisfies the following equation, it is called the ARCH(\( p \)) model:

\[
y_t = \mu + \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 \tag{41}
\]

where \( p \) is a positive constant. We set \( \alpha_0 > 0, \alpha_i \geq 0 \ (i = 1, ..., p) \), and \( \epsilon_t \) is standard Gaussian white noise, \( \epsilon_t \overset{\text{i.i.d.}}{\sim} N(0, 1) \). The ARCH model stands for Auto-Regressive Conditional Heteroscedastic model. (Shiraishi, 2022, and Shumway and Stoffer, 2017)

Second, we define a more general ARCH model, which is called a general ARCH model.

**Definition 19 (GARCH Model).** When \( \{y_t\} \) satisfies the following equation, it is called the GARCH(\( p, q \)) model:

\[
y_t = \mu + \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{42}
\]

where \( p \) and \( q \) are positive constants. We set \( \alpha_0 > 0, \alpha_i \geq 0 \ (i = 1, ..., p) \), \( \beta_j \geq 0 \ (j = 1, ..., q) \) and \( \epsilon_t \) is standard Gaussian white noise, \( \epsilon_t \overset{\text{i.i.d.}}{\sim} N(0, 1) \).

The GARCH model stands for the Generalized Auto-Regressive Conditional Heteroscedastic model. (Shiraishi, 2022, and Shumway and Stoffer, 2017)

Then we define the third model to analyze the exchange rate in this thesis.

**Definition 20 (Fractional Differencing Model).** Let us consider a time series \( \{y_t\} \) that satisfies the below equation

\[
\epsilon_t = (1 - B)^d y_t = \sum_{j=0}^{\infty} \pi_j B^j y_t = \sum_{j=0}^{\infty} \pi_j y_{t-j}, \quad d > -1 \tag{43}
\]

where \( \epsilon_t \) still denotes white noise with variance \( \sigma^2 \) and \( B \) is a back shift operator which can be expressed as \( B^j y_t = y_{t-j} \ (j \in \mathbb{N}) \). This model is called the fractional differencing model. In the equation (43), \( (1 - B)^d \) is the binomial expansion as

\[
(1 - B)^d = 1 - \binom{d}{1} B + \binom{d}{2} B^2 - \binom{d}{3} B^3 + ... = \sum_{j=0}^{\infty} \pi_j B^j \tag{44}
\]

where
\[ \pi_j = \frac{\Gamma(j - d)}{\Gamma(j + 1) \Gamma(-d)} \]  

with \( \Gamma(x + 1) = x\Gamma(x) \) being the gamma function. The gamma function is described as

\[ \Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx. \]

On the other hand, changing \( d > -1 \) to \( d < 1 \) yields

\[ y_t = (1 - B)^{-d} \epsilon_t = \sum_{j=0}^{\infty} \psi_j B^j \epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad d < 1 \]

where

\[ \psi_j = \frac{\Gamma(j + d)}{\Gamma(j + 1) \Gamma(d)}. \]

(Shumway and Stoffer, 2017)

3 Result

As in the introduction section, the above models are used to analyze the exchange rate between the Japanese Yen and the U.S. Dollar. The source data frame is taken from Dollar Yen Exchange Rate, 2023. It is recorded per day except for Japanese weekends and the period is from January 4th, 1971 to August 22nd, 2023, but the period before February 1st, 1973 is intentionally deleted. It is because that the worth of 1 dollar were fixed as 308 yen due to the government policy called Smithsonian system from December in 1971 to January in 1973. (Sumitomo Mitsui DS asset Management Company, 2022)

The edited data frame has 13069 data and the parameter of that is only the price of Japanese Yen against 1 U.S. Dollar. We can see the data frame and the plot that we use in Figure 1.

3.1 Investigations of Data Properties

First of all, we check the ACF and the PACF to understand the features of the three different data frames. All of them show that the ACFs tail off and the PACFs cut off after lag 1. However, when we take a difference for the exchange data, its ACF and its PACF are small enough to approximate with 0. That means that the exchange data is white noise and supports that we can fix \( d = 1 \).

In general, it is common to take a difference for log-transformed data as \( \log y_t - \log y_{t-1} \). And it can be transformed as
\[
\log y_t - \log y_{t-1} = \log \left[ 1 + \frac{y_t - y_{t-1}}{y_{t-1}} \right] \approx \frac{y_t - y_{t-1}}{y_{t-1}} =: x_t
\]

where \( x_t \) is small enough to be approximated. \( x_t \) is called an earning rate and it is an important target to be analyzed in the financial analysis. (Tanaka, 2006)

### 3.2 Result of ARIMA Model

Second, the exchange data is analyzed with the ARIMA model after preparing for the non-transformed data frame, the log-transformed data frame, and the square-root-transformed data frame. We will estimate 17 ARIMA models with \( d = 1 \). These models are made per one transformed data frame. Additionally, we can compare with the Ljung-Box statistic graphs of each ARIMA model with \( m = 20 \) fixed. To see them, all of the Ljung-Box statistic p values in \( m = 10 \) are above blue lines, so that test outcomes are not to reject the null hypothesis. It means that three time series with each transform are valid because the residuals of each model follow a white noise.

However, when we compare each residuals graph of three ARIMA models, one with log-transformed is similar to a normal distribution. So, we understand that the ARIMA model with log-transformed data frame is better than the other ARIMA models. By these investigations, we can also say that using the log-transformed data frame is the best way to analyze the exchange data.

We can see the ACF plots, the PACF plots, the histograms of each residual, Q-Q plots, the plots of the Ljung-Box statistics, and the graphs of the results. The first one is plotted by the non-transformed data frame, the second one is made by the log-transformed data frame, and the square-transformed data frame is used in the final one. RMSE values and MAE values in Table 4 are calculated after changing transformations into the normal form.

### 3.3 Result of GARCH Model and Fractional Differencing Model

As the 2nd model, the exchange data is analyzed with the GARCH model. We will use the log-transformed data because of the result with the ARIMA model and make nine GARCH \((p, q)\) models where \( 1 \leq p, q \leq 3 \). According to Table 4, there are few differences between the AIC, the BIC, the RMSE and the MAE values of them. So, we cannot judge which model is better. But both their RMSE and MSE values are smaller than ARIMA models’ ones. Hence we can think GARCH models are better than ARIMA models in this prediction. We understand that their residuals also have good properties by Figure 6 to Figure 9.

Finally, we can find the feature of the fractional differencing model as a 3rd model that the log-transformed data frame is used for the same reason as the
GARCH model. The AIC of this model is $-57139.31$. But both its RMSE and its MAE in Table 4 are the smallest values, so we understand that this model is the best model of the three types of models. However, when we see Figure 9, it is questionable that the residuals do not seem to be close to the normal distribution like ARIMA models and GARCH models and then those all ACF and PACF values are not near 0 where the lag $h$ is in $1 \leq h \leq 120$.

We can see result of all predictions in Figure 10 to 13.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\phi_1$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\phi_1$ s.e.</th>
<th>$\lambda_1$ s.e.</th>
<th>$\lambda_2$ s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA (1, 1, 2) with normal data</td>
<td>0.9409</td>
<td>-0.9048</td>
<td>-0.0185</td>
<td>0.0242</td>
<td>0.0257</td>
<td>0.0094</td>
</tr>
<tr>
<td>ARIMA (1, 1, 0) with log-data</td>
<td>0.0175</td>
<td>***</td>
<td>***</td>
<td>0.0087</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>ARIMA (1, 1, 0) with square-root-data</td>
<td>0.0274</td>
<td>***</td>
<td>***</td>
<td>0.0087</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

Table 1: Parameter estimates and standard errors (s.e.) of the ARIMA models
<table>
<thead>
<tr>
<th>Model</th>
<th>$\mu$</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH (1, 1)</td>
<td>4.539e-05</td>
<td>3.737e-07</td>
<td>8.569e-02</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH (2, 1)</td>
<td>4.408e-05</td>
<td>3.734e-07</td>
<td>8.553e-02</td>
<td>1.000e-08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH (2, 2)</td>
<td>5.248e-05</td>
<td>4.590e-07</td>
<td>1.071e-01</td>
<td>1.000e-08</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH (3, 1)</td>
<td>4.484e-05</td>
<td>3.733e-07</td>
<td>8.542e-02</td>
<td>1.000e-08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH (3, 2)</td>
<td>5.296e-05</td>
<td>4.581e-07</td>
<td>1.069e-01</td>
<td>1.000e-08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH (3, 3)</td>
<td>5.788e-05</td>
<td>4.703e-07</td>
<td>1.167e-01</td>
<td>1.000e-08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH (2, 3)</td>
<td>5.788e-05</td>
<td>4.703e-07</td>
<td>1.167e-01</td>
<td>1.000e-08</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>$\alpha_3$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH (1, 1)</td>
<td>***</td>
<td>9.126e-01</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>GARCH (2, 1)</td>
<td>***</td>
<td>9.128e-01</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>GARCH (2, 2)</td>
<td>***</td>
<td>5.449e-01</td>
<td>3.455e-01</td>
<td>***</td>
</tr>
<tr>
<td>GARCH (1, 2)</td>
<td>***</td>
<td>5.449e-01</td>
<td>3.455e-01</td>
<td>***</td>
</tr>
<tr>
<td>GARCH (3, 1)</td>
<td>1.000e-08</td>
<td>9.128e-01</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>GARCH (3, 2)</td>
<td>1.000e-08</td>
<td>5.480e-01</td>
<td>3.426e-01</td>
<td>***</td>
</tr>
<tr>
<td>GARCH (3, 3)</td>
<td>1.000e-08</td>
<td>6.109e-01</td>
<td>1.000e-08</td>
<td>2.702e-01</td>
</tr>
<tr>
<td>GARCH (2, 3)</td>
<td>***</td>
<td>6.110e-01</td>
<td>1.000e-08</td>
<td>2.702e-01</td>
</tr>
<tr>
<td>GARCH (1, 3)</td>
<td>***</td>
<td>6.110e-01</td>
<td>1.000e-08</td>
<td>2.702e-01</td>
</tr>
</tbody>
</table>

Table 2: Parameter estimates and standard errors (s.e.) of GARCH models
(a) ACF and PACF of the exchange data
(b) ACF and PACF of the exchange data after taking one difference

Figure 2: ACF plots and PACF plots

<table>
<thead>
<tr>
<th>Model</th>
<th>d estimate</th>
<th>d standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractional Differencing Model</td>
<td>0.4999542</td>
<td>0.0009797255</td>
</tr>
</tbody>
</table>

Table 3: The estimate and the standard error of the fractional differencing model

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA (1, 1, 2) with normal data</td>
<td>36589.93*</td>
<td>***</td>
<td>20.48155</td>
<td>15.2442</td>
</tr>
<tr>
<td>ARIMA (1, 1, 0) with log-data</td>
<td>-95046.95</td>
<td>***</td>
<td>20.25692</td>
<td>15.05076</td>
</tr>
<tr>
<td>ARIMA (1, 1, 0) with square-root-data</td>
<td>-48270.56*</td>
<td>***</td>
<td>20.26341</td>
<td>15.05635</td>
</tr>
<tr>
<td>GARCH (1, 1)</td>
<td>-7.461494</td>
<td>-7.459205</td>
<td>19.22849</td>
<td>14.20331</td>
</tr>
<tr>
<td>GARCH (2, 1)</td>
<td>-7.461507</td>
<td>-7.458645</td>
<td>19.31234</td>
<td>14.27345</td>
</tr>
<tr>
<td>GARCH (2, 2)</td>
<td>-7.463568</td>
<td>-7.460135</td>
<td>18.76358</td>
<td>13.8137</td>
</tr>
<tr>
<td>GARCH (1, 2)</td>
<td>-7.463721</td>
<td>-7.460860</td>
<td>18.76357</td>
<td>13.81369</td>
</tr>
<tr>
<td>GARCH (3, 1)</td>
<td>-7.461590</td>
<td>-7.458156</td>
<td>19.26542</td>
<td>14.23421</td>
</tr>
<tr>
<td>GARCH (3, 2)</td>
<td>-7.463589</td>
<td>-7.459583</td>
<td>18.73412</td>
<td>13.78987</td>
</tr>
<tr>
<td>GARCH (3, 3)</td>
<td>-7.465664</td>
<td>-7.461086</td>
<td>18.45708</td>
<td>13.55614</td>
</tr>
<tr>
<td>GARCH (2, 3)</td>
<td>-7.465817</td>
<td>-7.461812</td>
<td>18.45704</td>
<td>13.5561</td>
</tr>
<tr>
<td>GARCH (1, 3)</td>
<td>-7.465970</td>
<td>-7.462537</td>
<td>18.45704</td>
<td>13.5561</td>
</tr>
<tr>
<td>Fractional Differencing Model</td>
<td>-57139.31</td>
<td>***</td>
<td>15.00468</td>
<td>11.36811</td>
</tr>
</tbody>
</table>

Table 4: AIC, BIC, RMSE, and MAE values (values with "*" is not valid to be compared with others and values of "***" cannot be gained by functions on the programming)
Figure 3: Residuals’ plots of ARIMA (1, 1, 2) with unprogressed data
Figure 4: Residuals’ Plots of ARIMA (1, 1, 0) with log-transformed data
Figure 5: Residuals' Plots of ARIMA (1, 1, 0) with square-root-transformed data
Figure 6: Residuals of three GARCH models
Figure 7: Residuals of three GARCH models
Figure 8: Residuals of two GARCH models
Figure 9: Residuals of the GARCH model and the Fractional Differencing model
Figure 10: Prediction results of three ARIMA models and an GARCH model

Figure 11: Prediction results of four GARCH models
Figure 12: Prediction results of four GARCH models

Figure 13: Prediction results of the fractional differencing model
4 Conclusion

We understand that the fractional differencing model is the best model in this analysis. It is because that its RMSE and its MAE are the smallest values compared to the other values.

However, the models made in the analysis are not useful to predict a future rate in reality. In other words, the models do not follow transition of weak Japanese Yen against the U.S. Dollar although we use many past exchange data. So, we can guess that stationary of exchange data does not last long then we get better models if we predict an exchange rate by using data on the situation of the yen depreciation. Future study should be included to use a data with not long period but taken at finer time intervals.

5 Acknowledgements

I would like to thank Professor Rolf Larsson for supervising my work on this thesis. When I was at Uppsala University as an exchange student and after I came back home, he kindly taught me even though I did not know well about time series analysis when I started to work in Uppsala. I could not write quality contexts and complete this thesis in English without his supervision.

I am grateful to Professor Masaya Matsuura for supporting me with this thesis. When I was in trouble understanding some concepts, he advised me to comprehend them. He also supported me to study at Uppsala University safely last autumn and winter. I could not understand time series analysis well and complete this thesis without his support.

6 Appendix

6.1 Python Programming Code

Before using the below codes, the original data is edited. There are two things to be done. One is that top several lines with explanations of the data frame with letters are deleted with an editing application. The other one is that a space in front of the column name "value" is also deleted with the same means.

```python
import matplotlib.pyplot as plt
import pandas as pd

dy = pd.read_csv("dollar_yen.csv")
dy_0 = dy.drop(range(0,512))
dy_1 = dy_0["date"].str.split("/", n = 2, expand = True)
dy_2 = dy_1.set_axis(["year", "month", "day"], axis =1)
dy_3 = dy_2["year"].str.cat([dy_2["month"], dy_2["day"]])
dy_0["value").dtypes
dy_0["value"] = dy_0["value").astype(str)
```

27
dy_4 = dy_3.str.cat(dy_0["value"], sep = " ")
dy_5 = dy_4.str.split(" ", n =1, expand = True)
dy_6 = dy_5.set_axis(["time", "value"], axis =1)
dy_6.dtypes
dy_6["time"] = dy_6["time"].astype('int')
dy_6["value"] = dy_6["value"].astype('float')
dy_6.dtypes
dy_6.to_csv("exchange_data.csv", index= False)

plt.figure(figsize=(12,4))
plt.plot(dy_6["value"])  
plt.xlabel("time")
plt.ylabel("exchange rate")
plt.show()

6.2 R Programming Code

library(astsa)
library(xts)
library(fGarch)
library(fracdiff)
library(forecast)

yen_dollar <- read.csv("exchange_data.csv")
print(yen_dollar)
plot(yen_dollar$time, yen_dollar$value, type = "o")
acf2(yen_dollar$value, max.lag=100)

#The ARIMA model
yd_diff <- diff(yen_dollar$value)
plot(yd_diff)
acf2(yd_diff, max.lag=100)

yd.arima_00 <- arima(yen_dollar$value, order = c(0, 1, 0))
yd.arima_01 <- arima(yen_dollar$value, order = c(0, 1, 1))
yd.arima_10 <- arima(yen_dollar$value, order = c(1, 1, 0))
yd.arima_02 <- arima(yen_dollar$value, order = c(0, 1, 2))
yd.arima_11 <- arima(yen_dollar$value, order = c(1, 1, 1))
yd.arima_20 <- arima(yen_dollar$value, order = c(2, 1, 0))
yd.arima_03 <- arima(yen_dollar$value, order = c(0, 1, 3))
yd.arima_12 <- arima(yen_dollar$value, order = c(1, 1, 2))
yd.arima_21 <- arima(yen_dollar$value, order = c(2, 1, 1))
yd.arima_30 <- arima(yen_dollar$value, order = c(3, 1, 0))
yd.arima_04 <- arima(yen_dollar$value, order = c(0, 1, 4))
yd.arima_13 <- arima(yen_dollar$value, order = c(1, 1, 3))
yd.arima_22 <- arima(yen_dollar$value, order = c(2, 1, 2))
yd.arima_31 <- arima(yen_dollar$value, order = c(3, 1, 1))
yd.arima_40 <- arima(yen_dollar$value, order = c(4, 1, 0))
yd.arima_23 <- arima(yen_dollar$value, order = c(2, 1, 3))
yd.arima_32 <- arima(yen_dollar$value, order = c(3, 1, 2))

aic_yd.arima <- data.frame(x=c("yd.arima_00", "yd.arima_01", "yd.arima_10", "yd.arima_02", "yd.arima_11", "yd.arima_20", "yd.arima_03", "yd.arima_12", "yd.arima_21", "yd.arima_30", "yd.arima_04", "yd.arima_13", "yd.arima_22", "yd.arima_31", "yd.arima_40", "yd.arima_23", "yd.arima_32"), y=c(yd.arima_00$aic, yd.arima_01$aic, yd.arima_10$aic, yd.arima_02$aic, yd.arima_11$aic, yd.arima_20$aic, yd.arima_03$aic, yd.arima_12$aic, yd.arima_21$aic, yd.arima_30$aic, yd.arima_04$aic, yd.arima_13$aic, yd.arima_22$aic, yd.arima_31$aic, yd.arima_40$aic, yd.arima_23$aic, yd.arima_32$aic))

print(aic_yd.arima)

min(yd.arima_00$aic, yd.arima_01$aic, yd.arima_10$aic, yd.arima_02$aic, yd.arima_11$aic, yd.arima_20$aic, yd.arima_03$aic, yd.arima_12$aic, yd.arima_21$aic, yd.arima_30$aic, yd.arima_04$aic, yd.arima_13$aic, yd.arima_22$aic, yd.arima_31$aic, yd.arima_40$aic, yd.arima_23$aic, yd.arima_32$aic)

# Choose the ARIMA(1,1,2) model because it has the smallest AIC value in all AIC values of ARIMA models with normal data

yd.pr_12 <- predict(yd.arima_12, n.ahead=1000)
print(yd.arima_12)
print(yd.pr_12)
plot(yd.pr_12$pred)
length(yd.pr_12$pred)
acf2(yd.arima_12$residuals)

par(mfrow=c(2, 1))
hist(yd.arima_12$residuals)
qqnorm(yd.arima_12$residuals)
tsdia(yd.arima_12, 50)

yd_1_time <-
    yen_dollar$time[seq(1, length(yen_dollar$time)-1000)]
yd_1_value <-
    yen_dollar$value[seq(1, length(yen_dollar$value)-1000)]
yd_1 <- data.frame(yd_1_time, yd_1_value)

time_without_the_last <-
    seq(length(yen_dollar$time)-999, length(yen_dollar$time))

yd.arima_12_pred <- arima(yd_1$yd_1_value, order = c(1, 1, 2))
fore_arima_12 <- predict(yd.arima_12_pred, n.ahead=1000)

plot(time_without_the_last,
    yen_dollar$value[time_without_the_last],
    xlab='time', ylab='exchange rate', type="l")
title(main='ARIMA(1,1,2) exchange rate with non-transformed data')
lines(fore_arima_12$pred, col='red')

RMSE_ARIMA_normal <-
    sqrt(mean((yen_dollar$value[time_without_the_last]
             - fore_arima_12$pred)^2))
print(RMSE_ARIMA_normal)

MAE_ARIMA_normal <-
    mean(abs(yen_dollar$value[time_without_the_last]
             - fore_arima_12$pred))
print(MAE_ARIMA_normal)

# ARIMA models with both the log-transformed data frame and
# with square-transformed data frame are made by the same way of
# the ARIMA models with the normal data.

# The GARCH model
yd_diff.log <- diff(yen_dollar.log$value_log)
print(yd_diff.log)
plot(yd_diff.log)

acf2(yd_diff.log)

#GARCH(1, 1) model

summary(yd_11.g <- garchFit(~garch(1, 1), data=yd_diff.log))

acf2(yd_11.g@residuals)

par(mfrow=c(2, 1))
hist(yd_11.g@residuals)
qqnorm(yd_11.g@residuals)

yd_4_time <-
yen_dollar.log$time[seq(1, length(yen_dollar.log$time)-1000)]
yd_4_value <- yd_diff.log[seq(1, length(yd_diff.log)-999)]
yd_4 <- data.frame(yd_4_time, yd_4_value)

time_without_the_last_diff <-
  seq(length(yen_dollar.log$time)-999, length(yen_dollar.log$time)-1)
yd.garch_11_pred <-
garchFit(~garch(1, 1), data=yd_4$yd_4_value)

fore_garch_11 <- predict(yd.garch_11_pred, n.ahead=999)

plot(cumsum(yd_diff.log[time_without_the_last_diff]), xlab='time', ylab='accumulated exchange rate', type="l")
title(main='GARCH(1,1) exchange rate')
lines(cumsum(fore_garch_11$meanForecast), col='red')

vector_exp_999 <- rep(exp(1), length = 999)
original_1 <- cumsum(yd_diff.log[time_without_the_last_diff])
original_1 <-
  original_1 + yen_dollar.log$value_log[length(yen_dollar.log$time)-999]
prediction_1 <- cumsum(fore_garch_11$meanForecast)
prediction_1 <-
  prediction_1 + yen_dollar.log$value_log[length(yen_dollar.log$time)-999]

RMSE_GARCH_11 <-
sqrt(mean(((vector_exp_999)^(original_1))~(original_1))

print(RMSE_GARCH_11)

MAE_GARCH_11 <-
  mean(abs(((vector_exp_999)^(original_1))
    - (vector_exp_999)^(prediction_1)))
print(MAE_GARCH_11)

# The GARCH models are analyzed with the same way of
# the GARCH(1,1) model.

# The Fractional Differencing model
yd_12_time <-
  yen_dollar.log$time[seq(1, length(yen_dollar.log$time)-1000)]
yd_12_value <-
  yen_dollar.log$
  value_log[seq(1, length(yen_dollar.log$value_log)-1000)]
yd_12 <- data.frame(yd_12_time, yd_12_value)

time_without_the_last <-
  seq(length(yen_dollar.log$time)-999,
    length(yen_dollar.log$time))

tyd_ex = yd_12_value
tyd_ex.fd = fracdiff(yd_ex)
summary(yd_ex.fd)
pred.yd_ex.fd = forecast(yd_ex.fd, h=1000)

acf2(yd_ex.fd$residuals)
par(mfrow=c(2, 1))
hist(yd_ex.fd$residuals)
qqnorm(yd_ex.fd$residuals)

plot(yen_dollar.log$value_log[time_without_the_last],
  xlab='time', ylab='exchange rate', type="l")
title(main='Exchange rate of Fractional Differencing model')
lines(seq(0, 999), pred.yd_ex.fd$mean, col='red')

RMSE_Fractional <-
  sqrt(mean(((vector_exp)
    ^(yen_dollar.log$value_log[time_without_the_last])
    - (vector_exp)^(pred.yd_ex.fd$mean))^2)))
print(RMSE_Fractional)
MAE_Fractional <-
  mean(abs((vector_exp)
          "^(yen_dollar.log$value_log[time_without_the_last])
          - (vector_exp)"^(pred.yd.ex.fd$mean)))

print(MAE_Fractional)

7 References


(Source data) Dollar Yen Exchange Rate (USD JPY) - Historical Chart. URL: https://www.macrotrends.net/2550/dollar-yen-exchange-rate-historical-chart


Sumitomo Mitsui DS asset Management Company, 2022 (written in Japanese) URL: https://www.smd-am.co.jp/market/ichikawa/2022/10/irepo221021/


