IE313: Time Series Analysis
Problem Sets 2

Student Name: Nguyen Minh Duc
Student ID: 20202026

Problem 1. Let $\{Y_t\}$ be an $AR(2)$ process of the special form $Y_t = \phi_2 Y_{t-2} + e_t$. Find the ranges of values of ϕ_2 for which the process is stationary.

Solution.

Since $\{Y_t\}$ is an $AR(2)$ process with $\phi_1 = 0$, let's consider its characteristic equation

$$\begin{aligned} 1 - \phi_2 x^2 &= 0 \\ x^2 &= \frac{1}{\phi_2} \\ x &= \pm\frac{1}{\sqrt{\phi_2}}. \end{aligned}$$

The process is stationary if and only if $|x| > 1$, which is equivalent to

$$\begin{aligned} |x| = \left|\pm\frac{1}{\sqrt{\phi_2}}\right| = \left|\frac{1}{\sqrt{\phi_2}}\right| &> 1 \\ \frac{1}{\left|\sqrt{\phi_2}\right|} &> 1 \\ \left|\sqrt{\phi_2}\right| &< 1. \end{aligned}$$

If $\phi_2 \geq 0$, $\left|\sqrt{\phi_2}\right| = \sqrt{\phi_2} < 1 \implies 0 \leq \phi_2 < 1$.
Otherwise, when $\phi_2 < 0$, $\left|\sqrt{\phi_2}\right| = \left|i\sqrt{-\phi_2}\right| = \left|\sqrt{-\phi_2}\right| = \sqrt{-\phi_2} < 1 \implies -\phi_2 < 1 \implies 0 > \phi_2 > -1$.
Combining both cases, we obtained $-1 < \phi_2 < 1$. Therefore, the range of values of ϕ_2 for which the process is stationary is $(-1, 1)$.

Problem 2.

(a) Choose any time series data that you want to analyze. You may download it from the following websites.

- https://www.aihub.or.kr/
- https://www.kaggle.com/datasets
- https://research.google/tools/datasets/
- https://www.investing.com/

You should describe what the data is about, other details (e.g., time period, frequency), and why you chose it.

(b) Choose an appropriate deterministic trend for your data and estimate it. Then plot the trend with your data. Discuss whether the estimated trend fits your data or not.

(c) Perform residual analysis. Discuss the results.

Solution.

(a) Currency exchange rates play important roles in the global economy. Exchange rates can influence international trade, investment decisions, tourism, and even the prices of everyday goods. Currency analysis is able to provide insights to numerous economic factors such as interest rates, inflation, and geopolitical events, which leads to a better understanding of both local and global economic states. For both individuals and businesses, fluctuations in exchange rates can lead to significant financial implications. In particular, international students' financial status is highly dependent on such changes as they frequently have to exchange money from their home country to their current studying country. As an international student in Korea myself, I am interested in the currency exchange rates between Korean Won (KRW) and my country currency (Vietnamese Dong, or VND for short).

The data is obtained from kaggle/datasets/kurone02/krw-to-vnd-exchange-rate, which contains the daily currency exchange rate between KRW and VND from 2020 to 2022. The data consists of two columns:

- Date: The time that a data point is recorded, is stored in the form of `YYYY-MM-DD`.
- Close: The closing price of the corresponding day, stored in the form of a floating point number.

Please refer to Table 1 and Figure 1 for the preview of the data.

Date	Close
2020-01-01	20.04463892
2020-01-02	20.01987144
2020-01-03	19.87153069
2020-01-04	19.87443686
2020-01-05	19.87443686

Table 1: The first 5 data points of the data

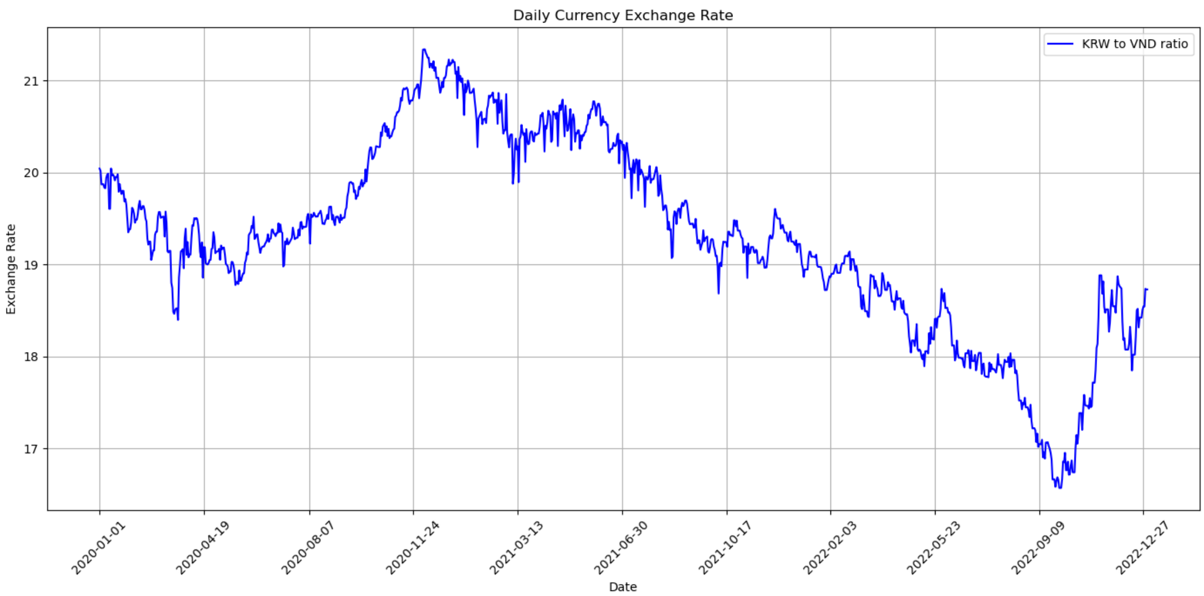

Figure 1: Time series plot of the data

Let's investigate the correlation between two consecutive days, which can be done by plotting a scatter plot. Figure 2 shows that that two consecutive days exhibit a strong positive correlation with each other. Doing the same thing with the monthly data gives similar conclusion. Please refer to Figures 3 and 4 for more details.

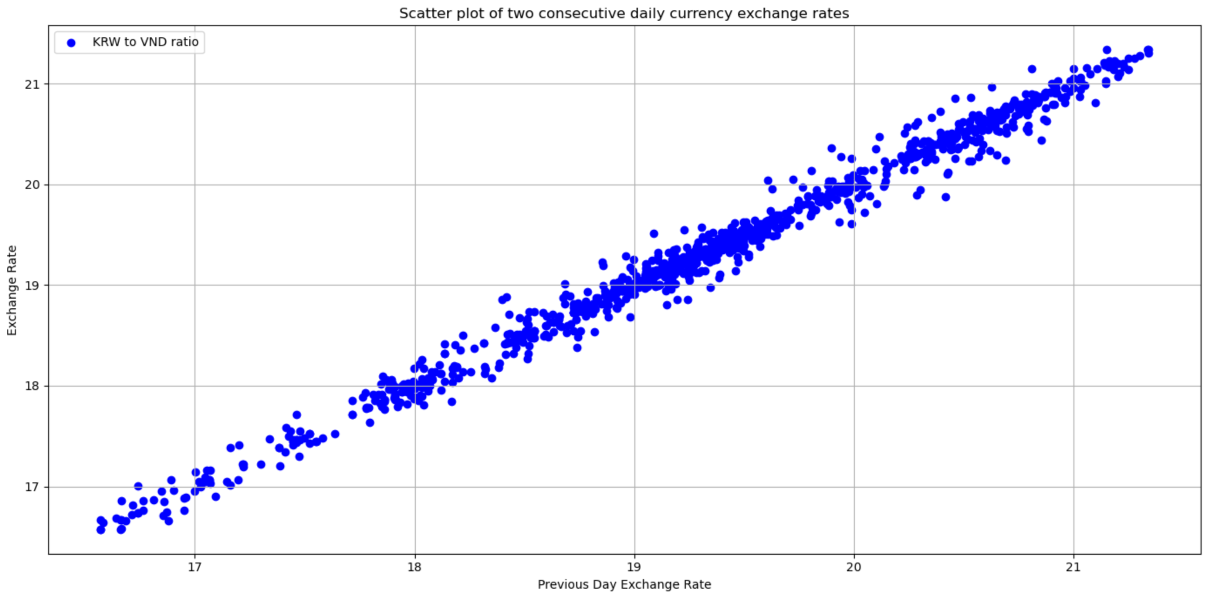

Figure 2: Two consecutive daily currency exchange rates

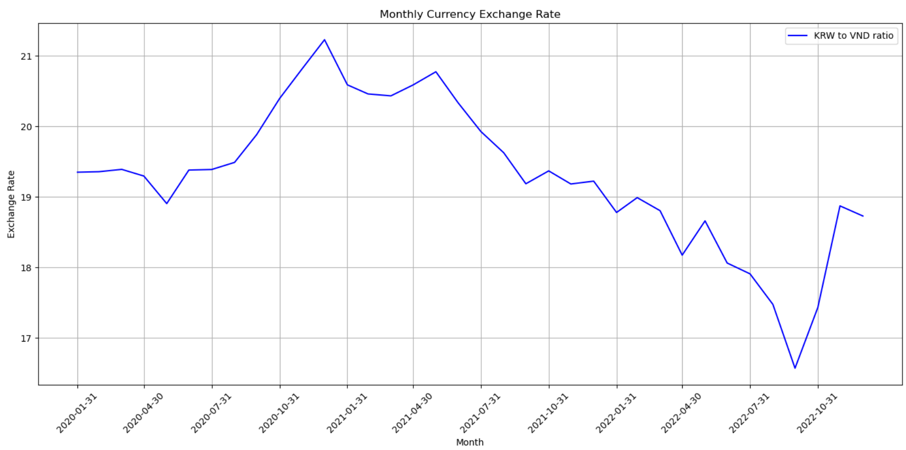

Figure 3: Monthly currency exchange rates

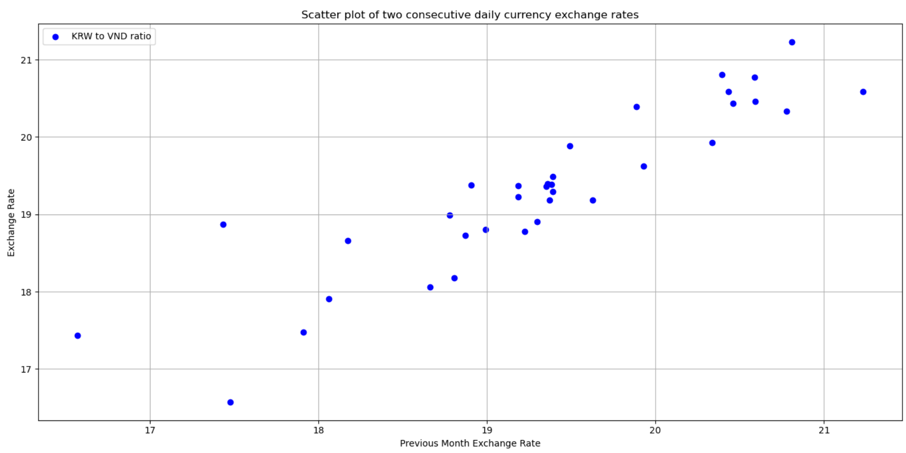

Figure 4: Two consecutive monthly currency exchange rates

(b) As one could observe, the data does not follow a simple linear or quadratic trend. Doing two simple regressions confirms the hypothesis, which is visualized in Figure 5. The linear line shows a decreasing trend in the overall shape of the data, and the quadratic curve follows quite closely to the middle part of the series. When dealing with complicated data like this, we can decompose it into sinusoidal waves, together with another polynomial, which, in this case, is a second-order polynomial,

$$f(x) = w_0 + w_1 x + w_2 x^2 + w_3 \sin(2\pi f_1 x) + w_4 \sin(2\pi f_2 x),$$

where f_1 and f_2 are two frequencies of the sinusoidal waves.

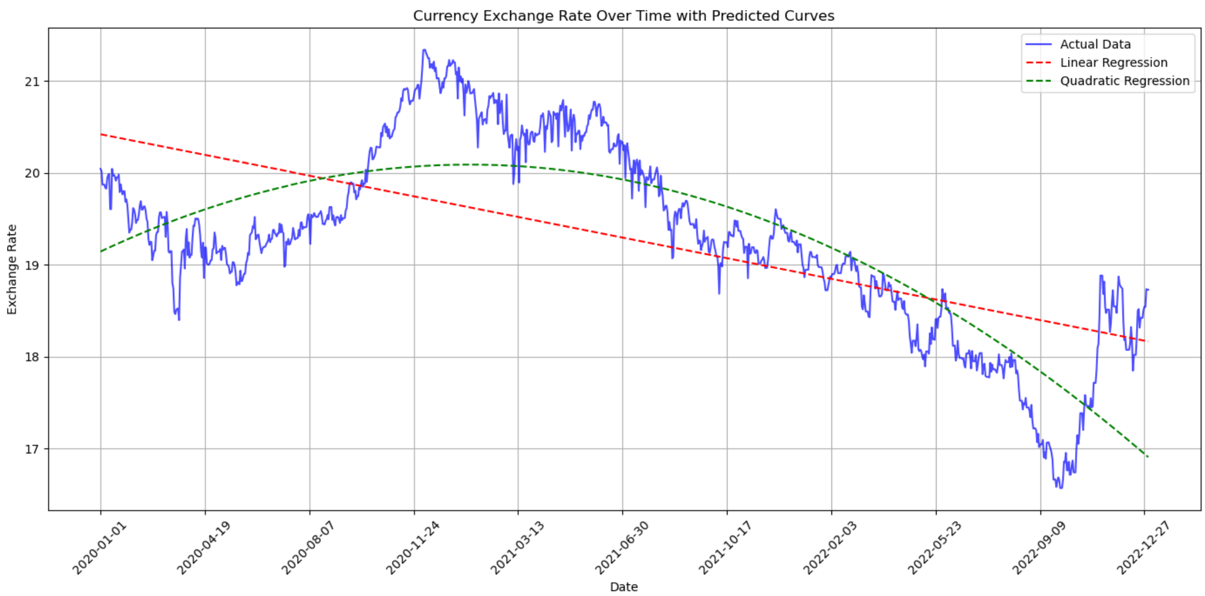

Figure 5: Simple linear and quadratic regression

This function can be rewritten in the form of matrix multiplication as

$$f(x) = \begin{bmatrix} w_0 & w_1 & w_2 & w_3 & w_4 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \\ \sin(2\pi f_1 x) \\ \sin(2\pi f_2 x) \end{bmatrix} = \mathbf{w}^T \mathbf{x}.$$

By doing a grid search, the following frequencies achieves best performance: $\begin{cases} f_1 &= \dfrac{(365 \times 0.21)^{-1}}{2\pi} \\ f_2 &= \dfrac{(365 \times 0.89)^{-1}}{2\pi} \end{cases}.$

Table 2 and Figure 6 summarize the results of a linear regression with $\mathbf{w}$ on the data.

Coefficient	**Estimate**	**Std. Error**	**t-value**	**$Pr(>\|t\|)$**
Intercept	20.2384	0.033	615.768	< 0.001
Time	-0.0794	0.001	-53.174	< 0.001
Time^2	7.558×10^{-5}	1.45×10^{-6}	52.051	< 0.001
$\sin(2\pi f_1 \text{Time})$	-0.4735	0.014	-33.136	< 0.001
$\sin(2\pi f_2 \text{Time})$	21.1348	0.375	56.371	< 0.001

Residual standard error	0.288 with 1090 degrees of freedom
Multiple R-Squared	0.925
Adjusted R-Squared	0.925
***F*-statistic**	3361. with 4 and 1090 df; p-value < 0.001

Table 2: The quadratic regression output

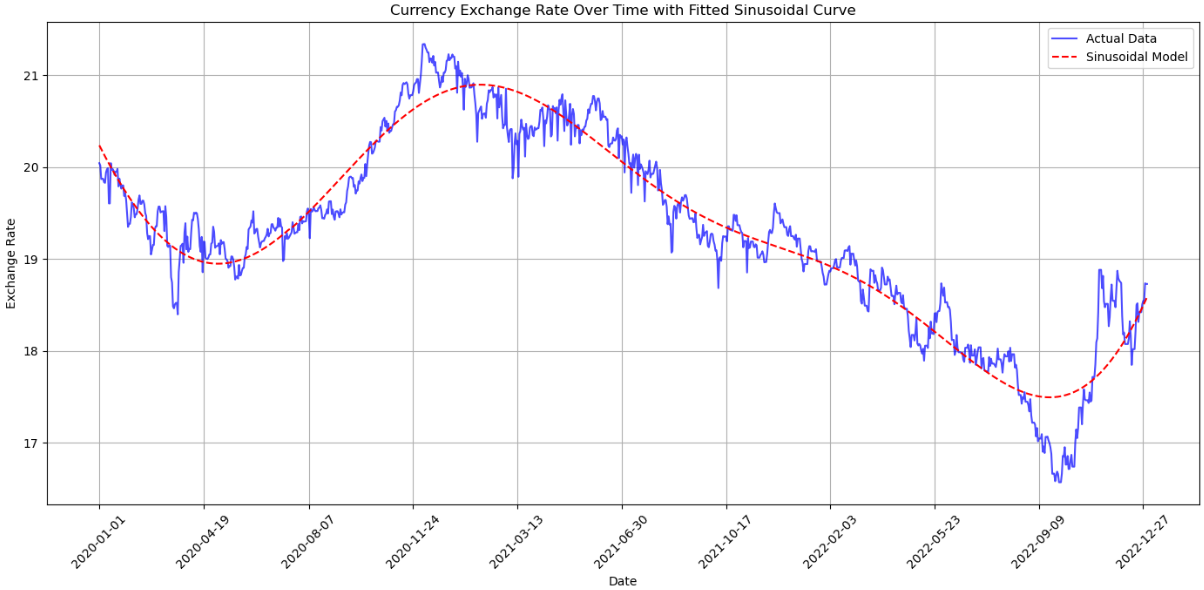

Figure 6: Fitted sinusoidal waves with second-order polynomial

The Residual standard error is quite small compared to the range of the data, hence, it indicates that our model is a good fit for the data. The R^2 score of this regression is 0.925, which is quite good. About 93% of the unobserved stochastic component in the wages series can be explained by the sinusoidal trend. The F-statistic is very high indicating that our results are statistically significant. As for the parameters, all of the standard deviations are small and all of the p-values are also extremely small, hence, the they are statistically significant. However, there's a period that does not follow our trend, which is around September, 2022. At that time, the price of the Korean Won was unstable and experienced a deep decrease, of which was not captured by our model.

(c) In this section, we will conduct a residual analysis to confirm that our model is a good fit to the data. Figure 7, 8 and 9 shows the plot for the standardized residuals, correlogram and side-by-side histogram and Q-Q plot respectively.

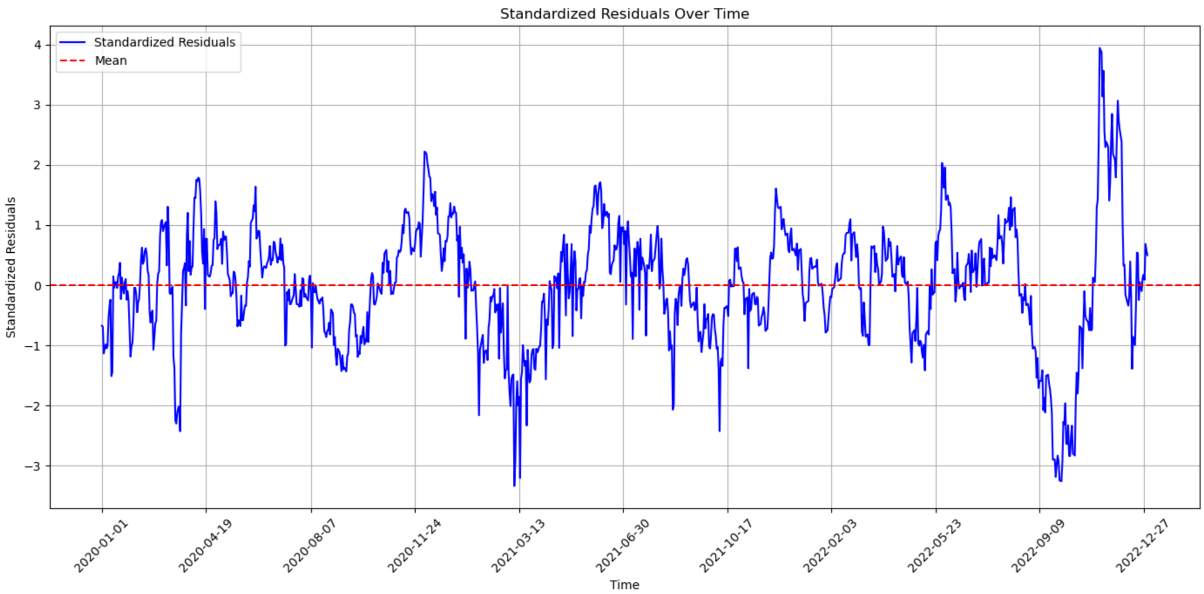

Figure 7: Time series plot for the standardized residuals

The plot looks stationary, the residuals are generally centered around zero, and there is no clear trend. However, there are maybe some recurring patterns left in the residuals.

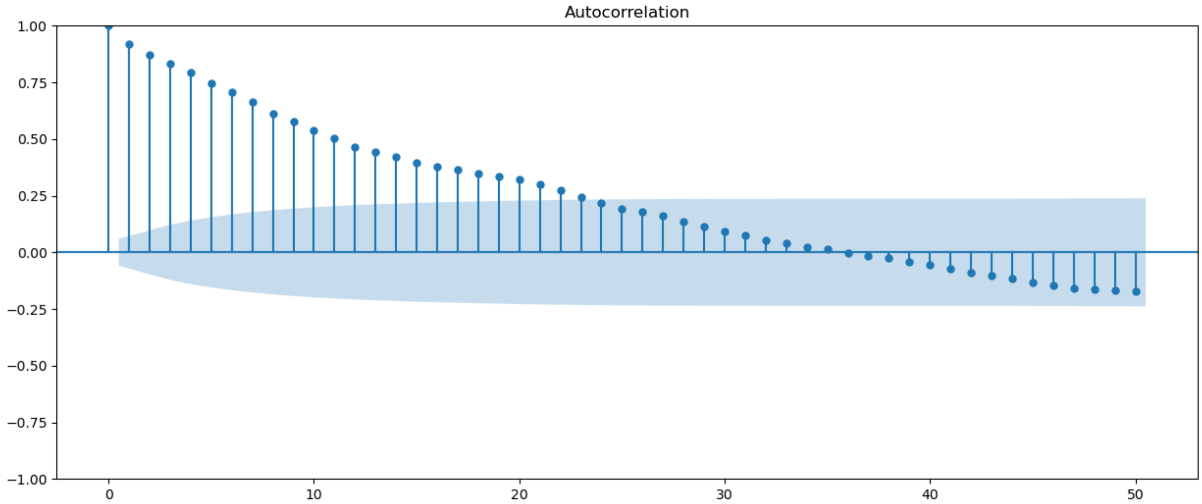

Figure 8: Plot for the sample autocorrelations for the standardized residuals

As for the autocorrelations, one could easily observe a decreasing trend in the magnitude of the correlation as the lag increases, i.e., there is a large correlation among neighboring residuals. Unfortunately, all of the lags that is less than 20 have their autocorrelation exceeds two standard errors, which is not the expected behavior of a white noise process, hence, we need to conduct more analysis on the spatial locality of the data.

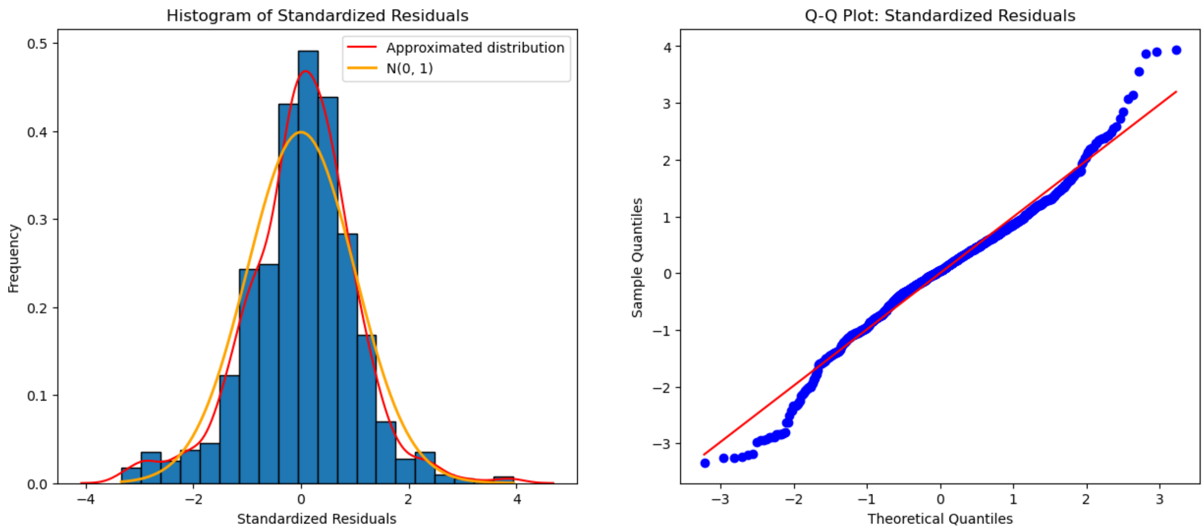

Figure 9: Histograms and normal probability plots for the standardized residuals

Looking at the histogram of the residuals, it resembles the bell curve of the normal distribution and has no significant skewness. However, it does have a much higher kurtosis. As for the Q-Q plot, most of the residuals are on the critical $y = x$ line, but some of them are off at the tails with a high degree of symmetry. To sum up the residual analysis, the residuals of our model looks stationary indicating that our suggested trend is fitted quite well. To fit to the data even better, it is a good idea to incorporate the local dependencies of the exchange rates in our model.