Time Series Analysis Project: Daily Euro-exchange rate

I. Introduction

Currency exchange rates play a critical role in global finance, international trade, and economic policy. In this project, our objective is to analyze historical foreign exchange rate data using Time Series Analysis (TSA) techniques. The data set consists of daily exchange rates for multiple currencies over a specified period, providing insight into market trends and volatility.

The data set includes the following columns: The Australian Dollar, Bulgarian Lev, Brazilian Real, Canadian Dollar, Swiss Franc, Chinese Yuan, Czech Koruna, Danish Krone, UK Pound Sterling, Japanese Yen, US Dollar, Euro, Norwegian Krone, Turkish Lira, and many others. These values represent the exchange rate of each currency against a base currency, the Euro. Some currencies in the dataset, such as the Cypriot Pound, Estonian Kroon, Greek Drachma, Lithuanian Litas, Latvian Lats, Maltese Lira, Slovenian Tolar and Slovak Koruna, no longer exist, having been replaced by the Euro over time.

The snippet of the dataset is shown below:

```
        Date [Australian dollar ] [Bulgarian lev ] [Brazilian real ]  \
0 2024-09-27                1.618           1.9558            6.0668
1 2024-09-26               1.6217           1.9558            6.0372
2 2024-09-25               1.6276           1.9558            6.1104
3 2024-09-24               1.6237           1.9558             6.137
4 2024-09-23               1.6274           1.9558            6.1976

  [Canadian dollar ] [Swiss franc ] [Chinese yuan renminbi ] [Cypriot pound ]  \
0             1.5036          0.942                    7.823              NaN
1             1.5025         0.9452                   7.8213              NaN
2             1.5044         0.9495                   7.8692              NaN
3             1.5033         0.9439                    7.829              NaN
4             1.5065         0.9448                   7.8438              NaN

  [Czech koruna ] [Danish krone ]  ... [Romanian leu ] [Russian rouble ]  \
0          25.158           7.457  ...          4.9764               NaN
1          25.159          7.4572  ...          4.9759               NaN
2          25.123          7.4575  ...          4.9760               NaN
3          25.155          7.4571  ...          4.9756               NaN
4          25.098          7.4581  ...          4.9742               NaN

  [Swedish krona ] [Singapore dollar ] [Slovenian tolar ] [Slovak koruna ]  \
0           11.273              1.4305                NaN              NaN
1             11.3              1.4328                NaN              NaN
2           11.311              1.4378                NaN              NaN
3          11.2935              1.4341                NaN              NaN
4           11.362              1.4357                NaN              NaN

  [Thai baht ] [Turkish lira ] [US dollar ]  [South African rand ]
0       36.135         38.1336       1.1158                19.1092
1       36.215         38.1022       1.1155                 19.238
2       36.599         38.2090       1.1194                19.1932
3       36.569         38.0064       1.1133                19.2869
4       36.654         37.9745       1.1119                19.3253

[5 rows x 41 columns]
Data shape: (6655, 41)
Date range: 1999-01-04 00:00:00 to 2024-09-27 00:00:00
```

Fig. 1. Data set insights

II. Understanding historical insights from correlation between two time series

In this project, we have looked at two time series: US dollars (USD) and Indian rupee (INR). in our data set, USD time series does not have missing values, and INR time series does not have data for the period 1999-01-04 to 2000-01-12, so we simply omit these records while analyzing the cross-correlation between these two time series:

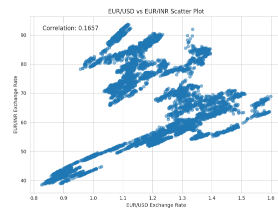

Fig. 2. Correlation using scatter-plot

Also, the top most correlated time series from our dataset are:

- US dollar & Chinese yuan renminbi: 1.0000
- Lithuanian litas & Hong Kong dollar: 0.9999
- US dollar & Hong Kong dollar: 0.9998
- Hong Kong dollar & Chinese yuan renminbi: 0.9998
- US dollar & Lithuanian litas: 0.9998

1) Observation: These currency pairs show very high correlation, meaning they changed in similar ways against the Euro. This is mainly because some currencies were pegged or managed closely, like the Hong Kong dollar and Chinese yuan being linked to the US dollar.

III. Seasonality :

We can see that the data does not show any clear seasonality. From the plot, there is no repeating pattern over a yearly period, and when we look at the autocorrelation function (ACF), there are no significant spikes at regular intervals. This means the data does not follow any seasonal cycle and does not have strong periodic behavior.

IV. Time-series analysis on USD time series

Fig. 3. USD Time Series

A. *Stationarity of time series*

We check if the mean of time series is constant and the autocorrelation function depends only on lag, which is the condition of the time series to be stationary.

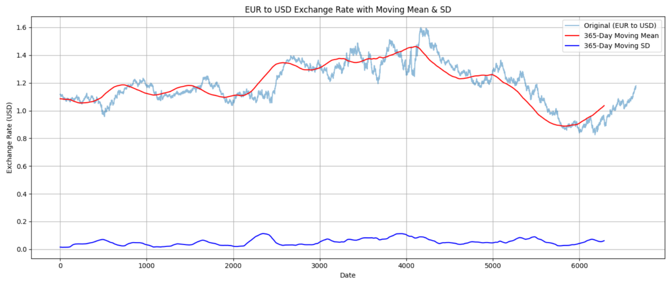

Fig. 4. 365-day Moving mean and standard deviation

We can see from the figure, that the mean is not constant and it changes with time, which violates the condition of the time series to be stationary.
Also, we can use ADF test, which is a statistical method used to determine whether a time series is stationary or non-stationary:
ADF Test: The Augmented Dickey-Fuller (ADF) test checks if a time series is stationary. A low p-value(< 0.05) means the series is stationary; a high p-value means it is not. For our time series,
ADF Statistic: -1.87751
p-value: 0.34271
Critical Values:
  1%: -3.43134
  5%: -2.86198
  10%: -2.56700
Conclusion: Since the ADF statistic is higher than all critical values and the p-value is large, we fail to reject the null hypothesis. The series is likely **non-stationary**.
To make the time series non-stationary we can use transformation, also we can detrend the series.

B. *Detrending using moving average method*

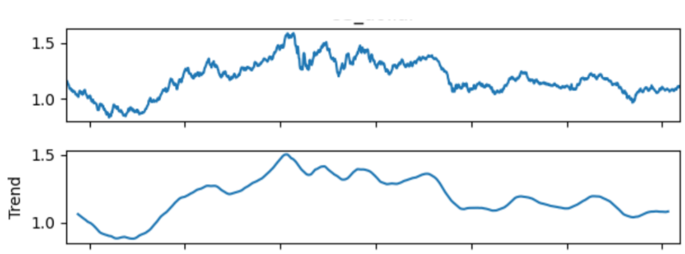

Fig. 5. Trend in USD time series

We have studied 3 methods of estimating trend : Moving average, regression and differencing. We can estimate trend and then remove it.

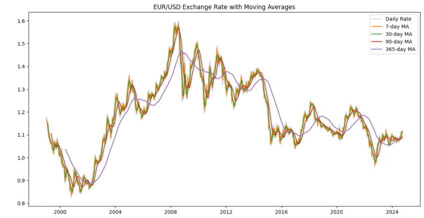

Fig. 6. Estimated trend using MA for different windows

We can see estimated trend from figure 8.
We can see that detrending of time series using moving average method for different windows make it stationary.

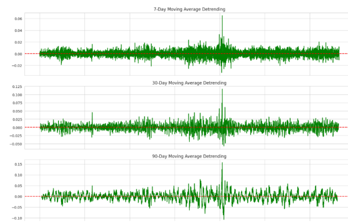

Fig. 7. Detrended time series for different windows

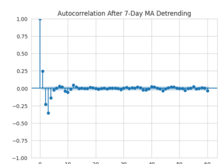

Fig. 8. ACF after 7 day MA detrending

C. *Regression using different orders*

We have tried to find trend assuming it is polynomial and we have used upto 8 degrees:

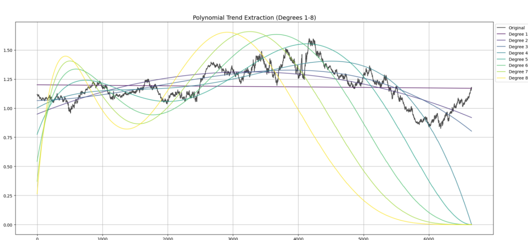

Fig. 9. Estimated polynomial trend using regression for different order

The results are shown below:

Polynomial Detrending Performance Comparison:

	Degree	ADF Statistic	p-value	Stationary	RMSE	BIC
0	1	-1.80338	0.37879	False	0.15541	-5820.40923
1	2	-2.92908	0.04207	True	0.10702	-10730.31196
2	3	-4.07707	0.00106	True	0.09753	-11946.68575
3	4	-6.70027	0.00000	True	0.20496	-2154.01093
4	5	-4.14350	0.00082	True	0.30801	3217.11182
5	6	-2.49577	0.11651	False	0.42301	7400.50330
6	7	-1.54972	0.50879	False	0.53237	10432.56288
7	8	-1.16489	0.68851	False	0.62544	12557.17377

Recommended polynomial degree: 2 (Lowest degree achieving stationarity)

Fig. 10. Stationarity and non-stationarity after detrending for different order

Observation: Polynomial detrending of degree 2 gives the best result, making the series stationary with the lowest RMSE (0.09753) and best BIC (-11946.69). For degrees above 4, the series becomes non-stationary again, likely due to overfitting and reintroducing trends.

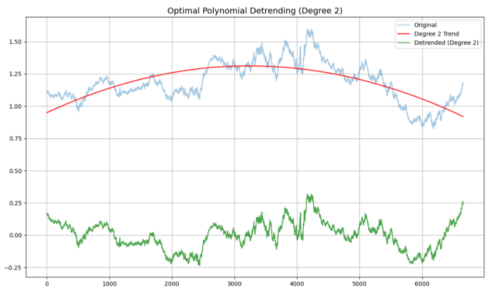

Fig. 11. Detrended stationary time series for best degree (2)

D. Differencing

Differencing is a fundamental technique in time series analysis used to render a non-stationary series stationary by removing trends or seasonality. By computing the difference between consecutive observations, it stabilizes the mean and variance. In our analysis, differencing by lag 1 (first order differencing) was applied.

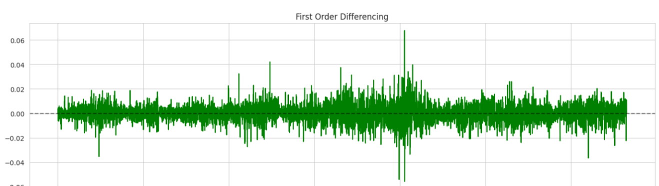

Fig. 12. time series after lag-1 differencing

First Differences:
ADF Statistic: -23.22191
p-value: 0.00000
Critical Values:
  1%: -3.43134
  5%: -2.86198
  10%: -2.56700

Conclusion: Since the ADF statistic is much lower than all critical values and the p-value is 0, we reject the null hypothesis. The first-differenced series is **stationary**.
Furthermore, autocorrelation function (ACF) plot demonstrated diminished autocorrelation, affirming stationarity

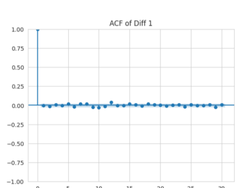

Fig. 13. ACF after 1st order differencing

V. Log Transformation Techniques for Time Series Stationarity

Logarithmic returns measure the relative change in an asset's value over time by using the natural logarithm of the ratio between the final and initial values. A kurtosis of 3 indicates that the distribution of returns has tail behavior similar to a normal distribution, meaning the likelihood of extreme events (like large gains or losses) is the same as in a normal distribution.

A. INR time series

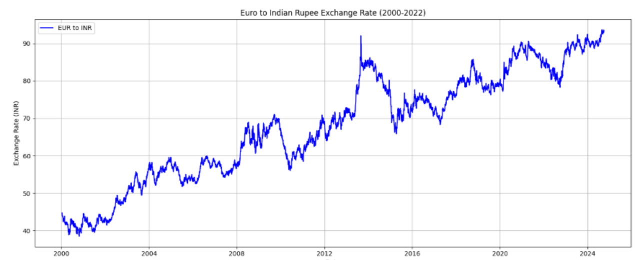

Fig. 14. INR time series

When we check for the stationarity of this time series, we find that it is not stationary.
We can check it using ADF test and also, by checking if the mean is constant or not

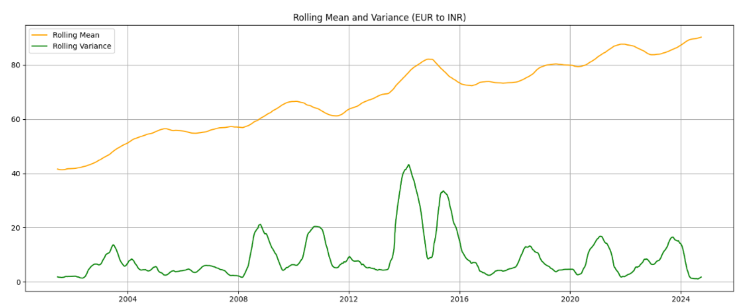

Fig. 15. Mean and standard deviation of INR time series

We can see that the mean is not constant, and it increases as time increases so it is function of time.

Observation: The INR time series becomes stationary after applying a log transformation. This is because the log scale compresses large values and reduces the impact of exponential growth or volatility. It helps stabilize the variance and diminish long-term trends, allowing the series to better satisfy the conditions of stationarity.

We can see it in figure 18.

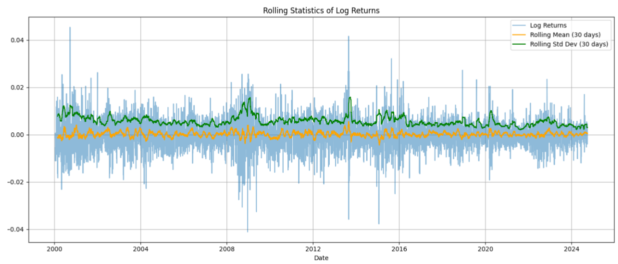

Fig. 16. Log transformation of INR time series

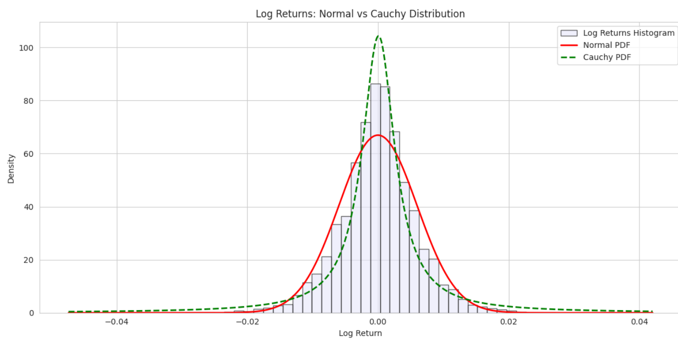

Fig. 17. Distribution of Log Returns with Fitted Normal and Cauchy Distribution

B. USD time series

The skewness close to zero (0.00486) suggests that the distribution of log returns is nearly symmetric, with the likelihood of extreme positive and negative returns being approximately equal. As we can see in Fig.19. On the other hand, the kurtosis of 6.15169 is significantly higher than 3, indicating that the distribution has heavier tails than a normal distribution. This suggests a higher probability of extreme events, such as large gains or losses, compared to a normal distribution, implying increased risk or volatility. A narrow spread indicates lower volatility, meaning that the exchange rates fluctuate less.

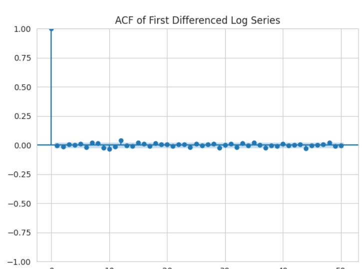

Fig. 18. Auto Corelation Function of Log Returns

We see that ACF plot for log returns drops to near zero quickly and stays there (within confidence bounds); this means that this log return series is essentially white noise - random and unpredictable.
‘

ADF Test:
ADF Statistic: -22.69031
p-value: 0.0000
Critical Values:
1%: -3.43141
5%: -2.86201
10%: -2.56702

Conclusion: Since the ADF statistic is much lower than all critical values and the p-value is 0, we reject the null hypothesis. The series is likely **stationary**.

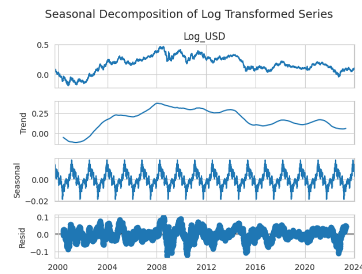

Fig. 19. Decomposition of log transformation into trend, seasonality and residual

VI. Conclusion :

The detailed analysis of the Euro to USD exchange rate helps us understand how this important currency pair changes over time. By studying historical data, we discovered long-term trends and seasonal patterns. We also used the log returns method to make the time series more stable and suitable for forecasting, by turning it into a stationary series. This approach gives us a clearer picture of the exchange rate behavior and helps build better prediction models.