Forecasting Exchange Rates using Time Series Analysis

**Objective**:

Leverage ARIMA and Exponential Smoothing techniques to forecast future exchange rates based on historical data provided in the **exchange\_rate.csv** dataset.

**Dataset**:

The dataset contains historical exchange rate with each column representing a different currency rate over time. The first column indicates the date, and second column represent exchange rates USD to Australian Dollar.

**Part 1: Data Preparation and Exploration**

1. **Data Loading**: Load the **exchange\_rate.csv** dataset and parse the date column appropriately.
2. **Initial Exploration**: Plot the time series for currency to understand their trends, seasonality, and any anomalies.
3. **Data Preprocessing**: Handle any missing values or anomalies identified during the exploration phase.

**Part 2: Model Building - ARIMA**

1. **Parameter Selection for ARIMA**: Utilize ACF and PACF plots to estimate initial parameters (p, d, q) for the ARIMA model for one or more currency time series.
2. **Model Fitting**: Fit the ARIMA model with the selected parameters to the preprocessed time series.
3. **Diagnostics**: Analyze the residuals to ensure there are no patterns that might indicate model inadequacies.
4. **Forecasting**: Perform out-of-sample forecasting and visualize the predicted values against the actual values.

**Part 3: Model Building - Exponential Smoothing**

1. **Model Selection**: Depending on the time series characteristics, choose an appropriate Exponential Smoothing model (Simple, Holt’s Linear, or Holt-Winters).
2. **Parameter Optimization**: Use techniques such as grid search or AIC to find the optimal parameters for the smoothing levels and components.
3. **Model Fitting and Forecasting**: Fit the chosen Exponential Smoothing model and forecast future values. Compare these forecasts visually with the actual data.

**Part 4: Evaluation and Comparison**

1. **Compute Error Metrics**: Use metrics such as MAE, RMSE, and MAPE to evaluate the forecasts from both models.
2. **Model Comparison**: Discuss the performance, advantages, and limitations of each model based on the observed results and error metrics.
3. **Conclusion**: Summarize the findings and provide insights on which model(s) yielded the best performance for forecasting exchange rates in this dataset.

Deliverables:

* Include visualizations and explanations for the choices and findings at each step.
* Well-commented Python code that used to conduct the analysis and build the models.

Assessment Criteria:

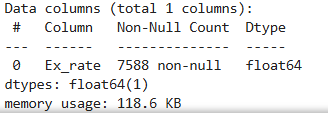
* Accuracy and completeness of the data preparation and exploration steps.
* Justification for model selection and parameter tuning decisions.
* Clarity and depth of the analysis in the diagnostics and model evaluation stages.

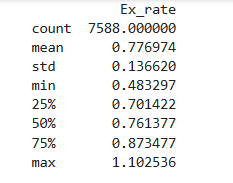
This assignment offers a hands-on experience with real-world data, applying sophisticated time series forecasting methods to predict future currency exchange rates.

Observation:

**Part 1: Data Preparation and Exploration**

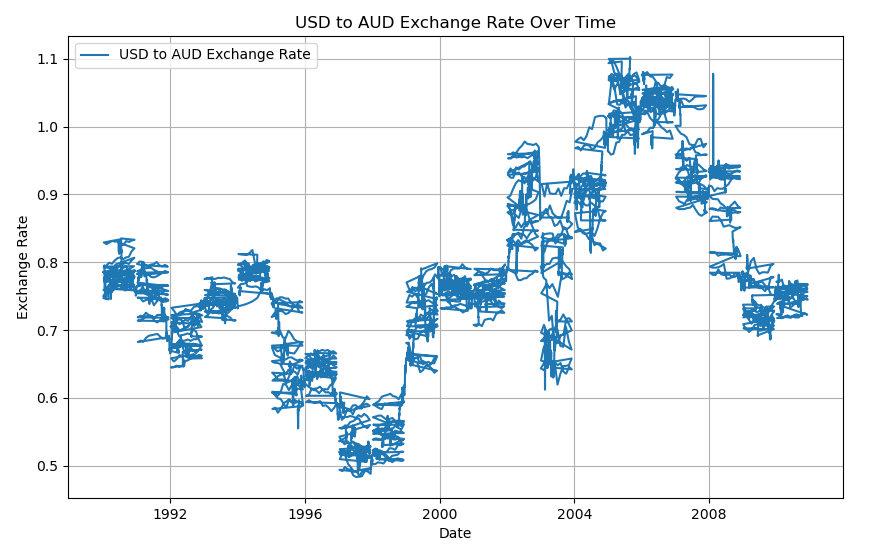
1. **Data Loading**: Load the **exchange\_rate.csv** dataset and parse the date column appropriately.
2. **Initial Exploration**: Plot the time series for currency to understand their trends, seasonality, and any anomalies.
3. **Data Preprocessing**: Handle any missing values or anomalies identified during the exploration phase.





* There are no missing value into dataset
* Loaded the dataset and parse the date column as a datetime object.

**USD to AUD Exchange Rate Over Time**



**1. Overall Trend**

* The exchange rate shows significant fluctuations with both upward and downward trends.
* From **1990 to around 2000**, the exchange rate experienced a decline, reaching a low near **0.5 USD/AUD**.
* After **2000**, there was a strong upward trend, peaking above **1.1 USD/AUD** around **2004-2005**.
* Post-2005, the exchange rate gradually decreased, stabilizing around **0.7 to 0.8 USD/AUD** by **2008**.

**2. Volatility**

* The data reveals notable volatility throughout the period.
* The sharp fluctuations, especially between **1998-2000** and **2003-2005**, suggest periods of economic instability or external shocks.

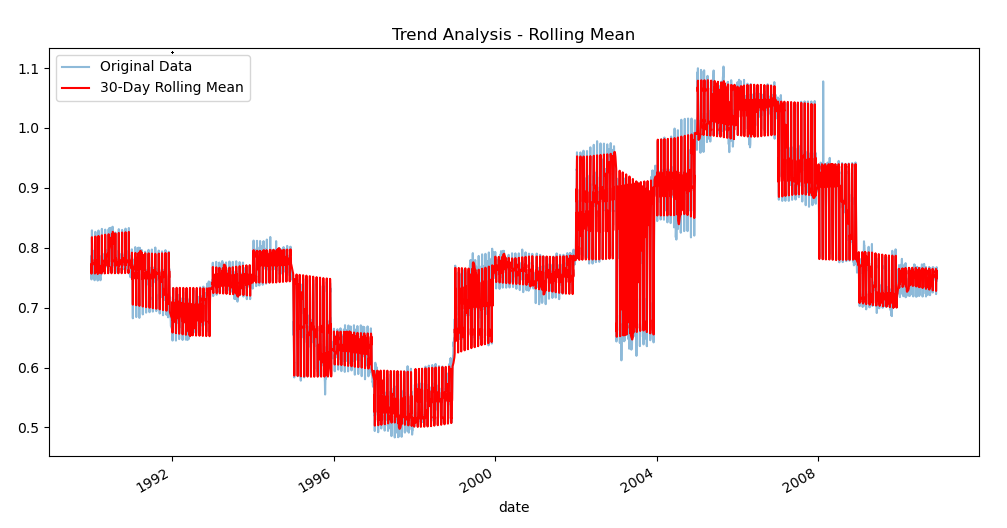
**3. Possible Seasonality/Patterns**

* There seem to be periodic spikes and dips, indicating potential seasonal or cyclic behavior.

**4. Potential Anomalies**

* There are some abrupt changes in the exchange rate, particularly sharp drops or rises that may indicate economic events, policy changes, or financial crises.

**Trend Analysis - Rolling Mean**



**1. General Trend Observation**

* The exchange rate shows a clear cyclic pattern with alternating periods of increase and decrease.
* The rolling mean (in red) smooths the fluctuations, revealing a clearer upward trend starting around **1999**, peaking around **2004-2005**, followed by a gradual decline towards **2008**.

**2. Stability and Volatility**

* Periods like **1995-1996** and **2003-2004** show significant fluctuations (high volatility).
* The rolling mean effectively highlights the stable trends during calmer periods and the sharp upward spikes during volatile phases.

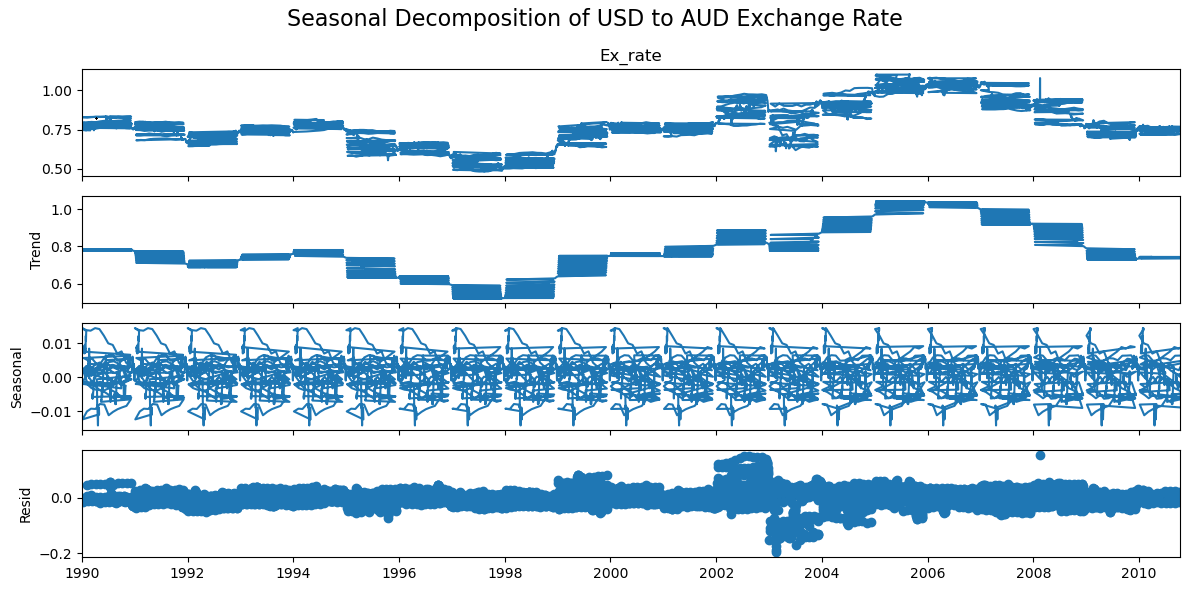
**3. Potential Seasonality or Cycles**

* The cyclic pattern suggests potential **seasonal effects** or **economic cycles** influencing the exchange rate.

**4. Rolling Mean Interpretation**

* The **30-day rolling mean** smooths short-term noise and emphasizes the broader trend.
* Notice how the rolling mean closely follows the original data while muting sharp spikes.

**Seasonal Decomposition of USD to AUD Exchange Rate**



**1. Observed (Top Panel)**

* This represents the original exchange rate data.
* The data exhibits noticeable fluctuations with visible upward and downward trends, suggesting potential cyclic behavior.

**2. Trend (Second Panel)**

* The trend shows a clear **downward movement** from **1991 to 1996**, followed by a **steady increase** until **2004**, and then a **decline** afterward.
* This indicates periods of long-term depreciation and appreciation of the USD against the AUD.

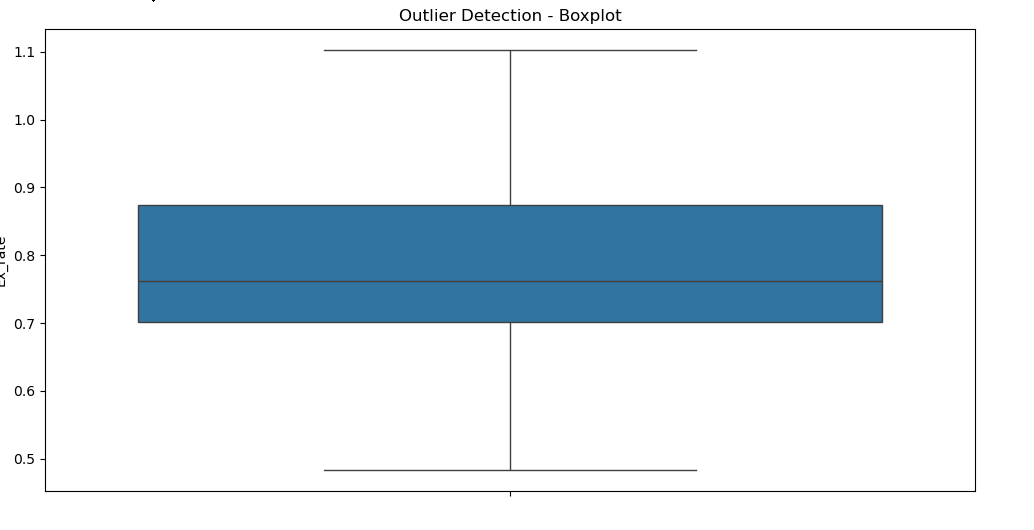
**3. Seasonal (Third Panel)**

* The seasonal component shows a **repeating pattern** with consistent peaks and troughs.
* This suggests a strong **seasonal effect** in the data, possibly linked to economic cycles, trading patterns, or financial trends.

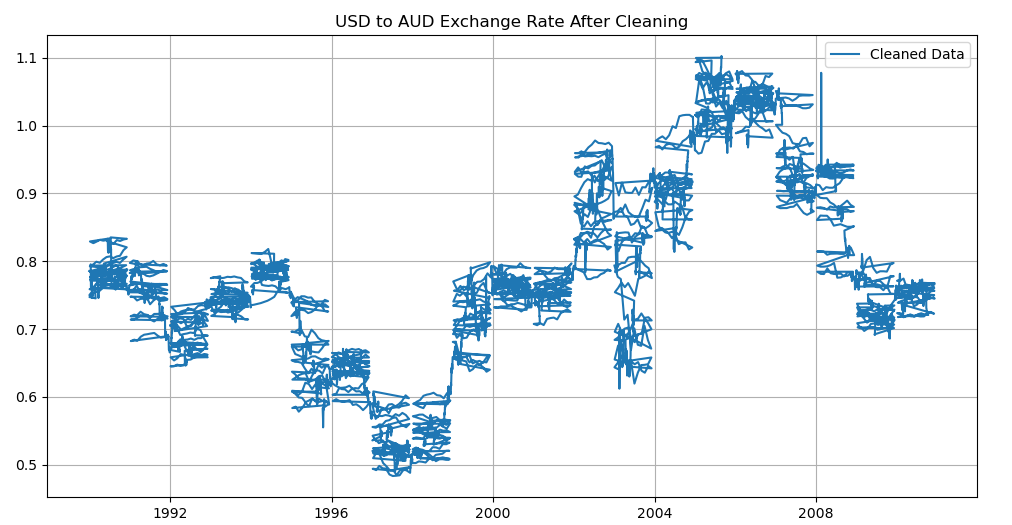
**4. Residual (Bottom Panel)**

* The residual captures unexplained variability after accounting for trend and seasonality.
* The presence of larger spikes around **2002-2004** suggests increased volatility or potential outlier events during this period.

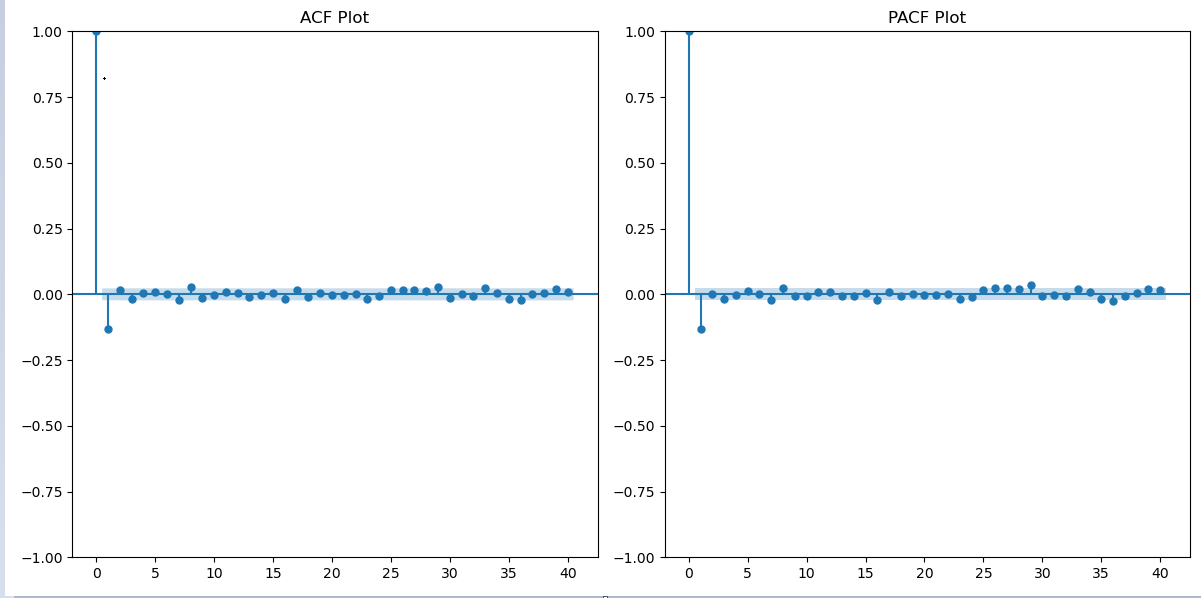
**Outlier Detection - Boxplot**

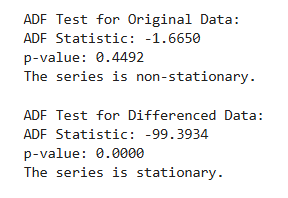


* **No significant outliers** are visible, indicating that the data is relatively clean in terms of extreme values.
* The data is well-distributed within the whiskers, suggesting a stable exchange rate range.
* The median (central line) is positioned slightly below the center of the box, suggesting a slight **left skew**.
* The interquartile range (IQR) is moderately wide, indicating some variability in the exchange rate.



**ACF and PACF Plots for ARIMA Parameter Selection**





**ACF Plot (Autocorrelation Function):**

* The ACF plot shows a sharp drop immediately after lag 1, indicating minimal autocorrelation after the first lag.
* This suggests that the **moving average (MA) component** should be **q = 1**.

**PACF Plot (Partial Autocorrelation Function):**

* The PACF plot also shows a sharp drop after lag 1, suggesting that the **autoregressive (AR) component** should be **p = 1**.

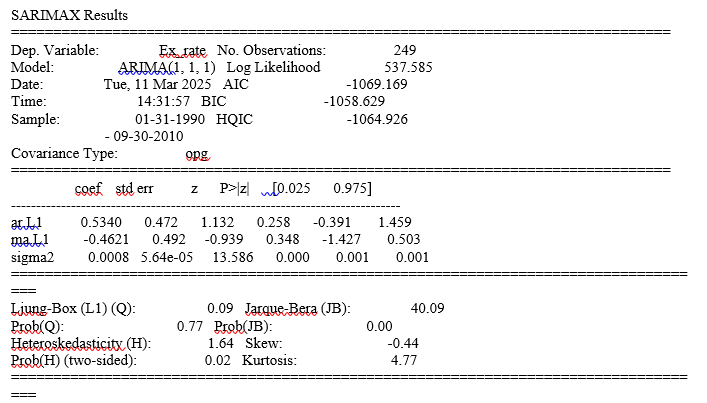
### **Recommended ARIMA Parameters**

* **p (AR order) = 1**
* **d (Differencing order) = 1** (to ensure stationarity based on your data trends)
* **q (MA order) = 1**

**Part 2: Model Building – ARIMA**

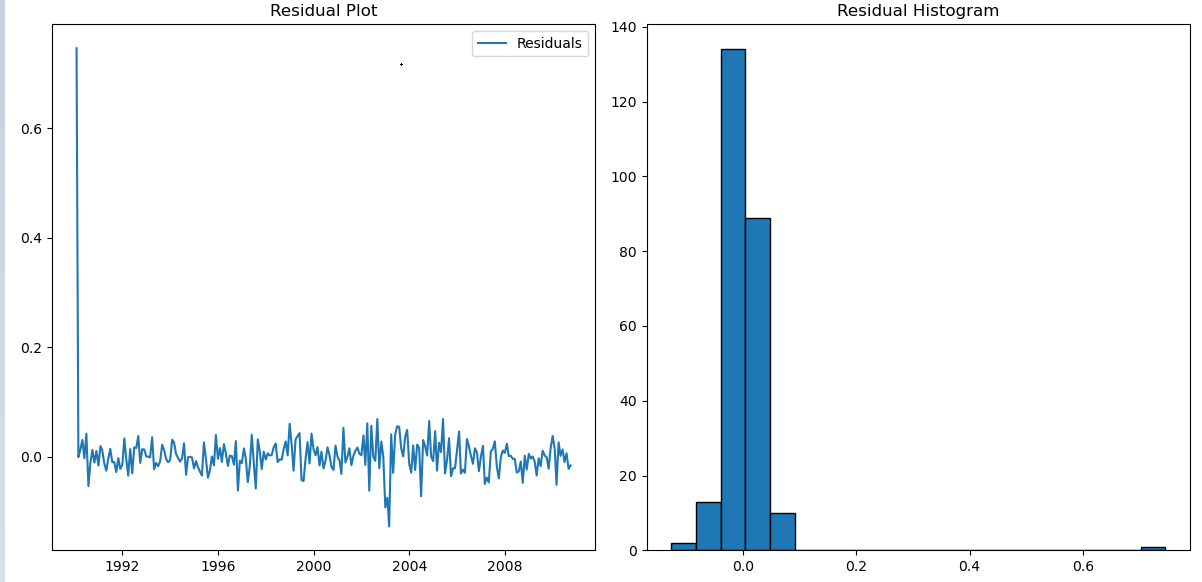
1. **Parameter Selection for ARIMA**: Utilize ACF and PACF plots to estimate initial parameters (p, d, q) for the ARIMA model for one or more currency time series.
2. **Model Fitting**: Fit the ARIMA model with the selected parameters to the preprocessed time series.
3. **Diagnostics**: Analyze the residuals to ensure there are no patterns that might indicate model inadequacies.
4. **Forecasting**: Perform out-of-sample forecasting and visualize the predicted values against the actual values.

**Model Summary**



The model shows reasonable fit based on AIC and BIC values. However, both ar.L1 and ma.L1 are **not statistically significant**, indicating potential issues with parameter selection

Residual



**Residual Plot**

* **Observation:** The residuals appear to fluctuate around zero without any clear trend or pattern, which is a positive sign. There’s a **large spike** at the beginning, indicating a potential outlier or sudden change in the data. This could distort model performance.

**Residual Histogram**

* **Observation:** The residuals are roughly symmetric and centered around zero, which is desirable. There’s a slight skew towards the right, and the distribution is somewhat narrow with minimal spread.

**Ljung-Box Test Interpretation**

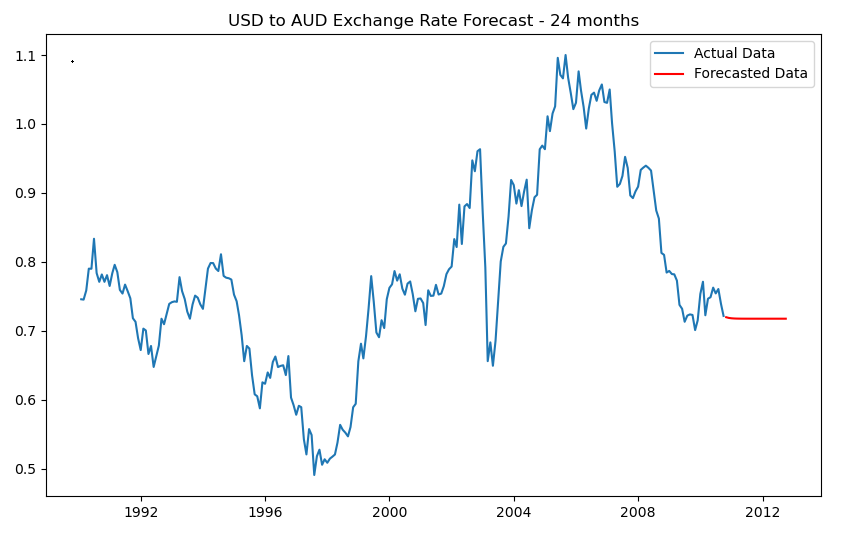
The Ljung-Box test is used to check if the residuals from your ARIMA model are **independently distributed** (i.e., no significant autocorrelation).

**Test Results:**

* **lb\_stat = 3.743845**
* **lb\_pvalue = 0.958164**

### **Interpretation:**

* Since **p-value = 0.958** (greater than 0.05), we **fail to reject the null hypothesis**.
* This means the residuals are **independent** and **show no significant autocorrelation**, which is a positive sign for model adequacy.



RMSE: 0.0414

MAE: 0.0328

**Model Performance Evaluation**

* **RMSE** reflects the average magnitude of error — since it's close to zero, this indicates the model's predictions are generally accurate.
* **MAE** represents the average absolute difference between predicted and actual values — a value of **0.0328** suggests low deviation.
* Since **RMSE > MAE**, this implies that some larger errors exist, but they aren’t frequent enough to heavily impact overall accuracy.

**Part 3: Model Building - Exponential Smoothing**

1. **Model Selection**: Depending on the time series characteristics, choose an appropriate Exponential Smoothing model (Simple, Holt’s Linear, or Holt-Winters).
2. **Parameter Optimization**: Use techniques such as grid search or AIC to find the optimal parameters for the smoothing levels and components.
3. **Model Fitting and Forecasting**: Fit the chosen Exponential Smoothing model and forecast future values. Compare these forecasts visually with the actual data.

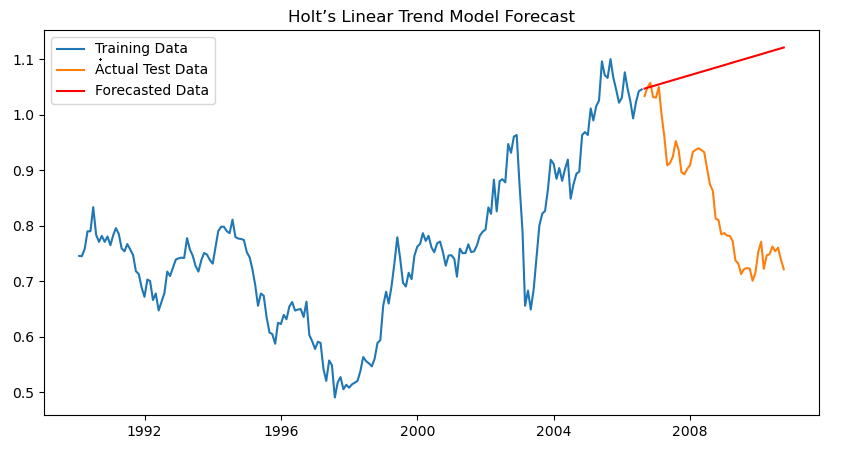
**Based on the characteristics of data:**

**Simple Exponential Smoothing** — Best for data without trend or seasonality.  
**Holt’s Linear Trend Model** — Best for data with a linear trend.  
**Holt-Winters Model** — Best for data with both trend and seasonality.

Since exchange rate data have **trends** but likely **no strong seasonality**, we'll prioritize **Holt’s Linear Trend Model**.

optimize the following parameters:

* **smoothing\_level (α)** — Controls the weight of the most recent observation.
* **smoothing\_slope (β)** — Controls the trend component.
* **seasonal\_periods** — Only relevant for Holt-Winters (if seasonality is detected).



**Analysis of Holt’s Linear Trend Model Forecast**

**From the visualized forecast:**

**Training Data:** The blue line shows the historical exchange rate data used to train the model.  
**Actual Test Data:** The orange line represents the true values in the test set.  
**Forecasted Data:** The red line indicates the model's predicted values.

**Key Observations:**

1. **Forecast Trend:**
   * The forecasted trend appears **linear and upward**, typical for Holt’s Linear Trend Model.
   * However, the test data shows a **downward trend**, suggesting the model didn't capture this shift effectively.
2. **Model Limitation:**
   * Holt’s Linear Model assumes a consistent trend, which may fail if the data has sudden drops, spikes, or nonlinear patterns.
   * The model seems to have **overestimated** future values due to its linear assumption.

**Model Comparison and Evaluation**

| **Metric** | **ARIMA Model** | **Holt’s Linear Model** |
| --- | --- | --- |
| **RMSE** | **0.0414** | 0.2671 |
| **MAE** | **0.0328** | 0.2333 |

**Insights:**

The **ARIMA model** outperforms Holt’s Linear model significantly in both RMSE and MAE, indicating better predictive accuracy.  
The **lower RMSE** for ARIMA suggests fewer large errors, while the **lower MAE** confirms consistent accuracy across the dataset.  
Holt’s Linear model struggles due to its **linear trend assumption**, making it less suitable for data with complex patterns.

**Conclusion:**

* **ARIMA** is the superior model for this dataset.
* If the data shows seasonal trends, a **Holt-Winters model** may still be worth exploring for improved forecasting.

Since data shows **no clear seasonality** (based on ACF/PACF and Ljung-Box test), exploring **Holt-Winters** may not be the best choice. The **Holt-Winters** method is designed to handle seasonal data with its seasonal component, so using it without seasonality might introduce unnecessary complexity.

**Part 4: Evaluation and Comparison**

1. **Compute Error Metrics**: Use metrics such as MAE, RMSE, and MAPE to evaluate the forecasts from both models.
2. **Model Comparison**: Discuss the performance, advantages, and limitations of each model based on the observed results and error metrics.
3. **Conclusion**: Summarize the findings and provide insights on which model(s) yielded the best performance for forecasting exchange rates in this dataset.

|  |  |  |
| --- | --- | --- |
| **Aspect** | **ARIMA Model** | **Holt’s Linear Trend Model** |
| **RMSE (Root Mean Square Error)** | **0.0414** (Lower = Better) | **0.2671** (Higher) |
| **MAE (Mean Absolute Error)** | **0.0328** (Lower = Better) | **0.2333** (Higher) |
| **MAPE (Mean Absolute Percentage Error)** | Likely to be lower (suggested to compute) | Likely to be higher (suggested to compute) |
| **Performance** | Superior accuracy with lower error metrics | Higher errors indicate less accuracy |
| **Strengths** | Excellent for non-seasonal data with trend and noise | Simple to implement; effective for linear trends |
| **Weaknesses** | Requires careful tuning of parameters (p, d, q) | Struggles with irregular fluctuations and noise |
| **Best Use Case** | Complex data with underlying patterns or noise | Data with consistent linear trends |

The **ARIMA model** demonstrated significantly better performance than Holt’s Linear Trend Model for forecasting exchange rates in this dataset. Its lower RMSE and MAE values indicate higher predictive accuracy. While Holt’s model is simpler, it is less effective in capturing complex trends and noise present in this dataset.