Project Report

École Supérieure de l'Économie Numérique

المدرسة العليا للاقتصاد الرقمي

Time Series Forecasting of the Exchange Rate Using ARIMA

AZIZ AISSI 3rd YEAR STUDENT BUSINESS INTELLIGENCE

EFFECTUÉ le 5 Fevrier 2026

TABLE OF CONTENTS

1.Introduction

Statistics plays a crucial role in understanding uncertainty and variability in real world phenomena. The objective of this project is to implement key concepts of inferential statistics studied during the course and apply them to a concrete dataset.

The main goals of this project are:

- To apply descriptive and inferential statistical methods
- To formulate and test statistical hypotheses
- To interpret results in a clear and meaningful way
- To demonstrate a solid understanding of statistical reasoning

2. Dataset Description

Dataset : Foreign Exchange Rates (2020–2025) — USD/EUR

Source : [Kaggle — Foreign Exchange Rates Datasets from 2020 to 2025 (by Sriharshithabattula)]

2)1.Dataset Overview

- A daily historical foreign exchange (forex) rates dataset from 2020 up to 2025.
- It includes exchange rates for major currency pairs:
  EUR/USD (Euro to US Dollar)
  GBP/USD (British Pound to US Dollar)
  USD/JPY (US Dollar to Japanese Yen)
  We will only use EUR/USD in the test
- The data is in a CSV file, small in size (~90 kB).

2)2.Why This Dataset Is Very Suitable for The Project ?

1. Strong Financial Relevance
- Exchange rates are a classic financial time series used widely for forecasting and econometric modeling in finance.
- Being daily data, it gives you a rich sequence of observations perfect for time series analysis.

2. Perfect for Time Series Modelling (ARIMA/ARMA)
- 5 years to capture trends, cycles, and volatility.
- Daily frequency
- realistic and complex time series on which to:
  - test stationarity
  - choose AR/I/MA parameters
  - analyze residuals
  - forecast with prediction intervals

2)3.Data Quality Verification

	Date	1.2 Close	1.2 High	1.2 Low	1.2 Open	ABC currency_pair
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
1	1/1/2020	1.122082591	1.122838497	1.115946889	1.122082591	EUR/USD
2	1/2/2020	1.122082591	1.122712493	1.116682053	1.121893764	EUR/USD
3	1/3/2020	1.11714375	1.11806798	1.112569809	1.117081285	EUR/USD
4	1/6/2020	1.116196036	1.120824933	1.115809917	1.116245866	EUR/USD
5	1/7/2020	1.119799495	1.119946241	1.113486528	1.119582534	EUR/USD

Prior to any statistical modeling, a data quality assessment was conducted using Power Query in Power BI.
This step aimed to ensure the reliability of the time series used for forecasting.

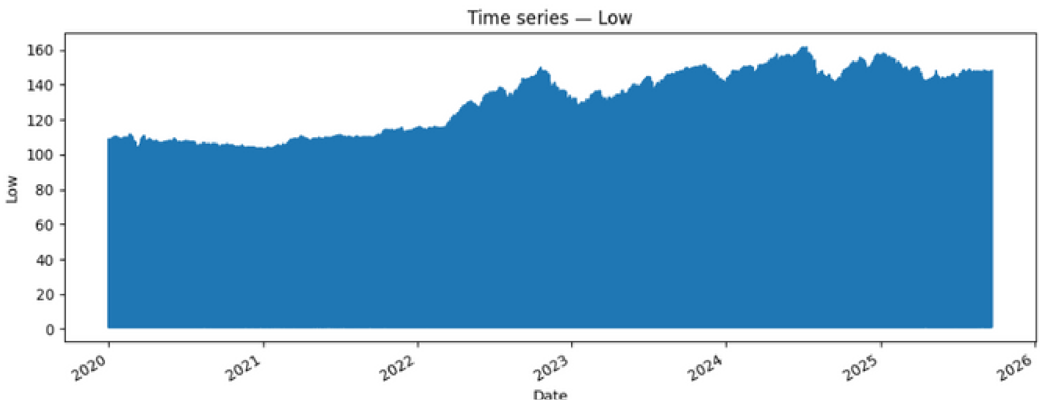
1.The Power Query quality indicators show that:

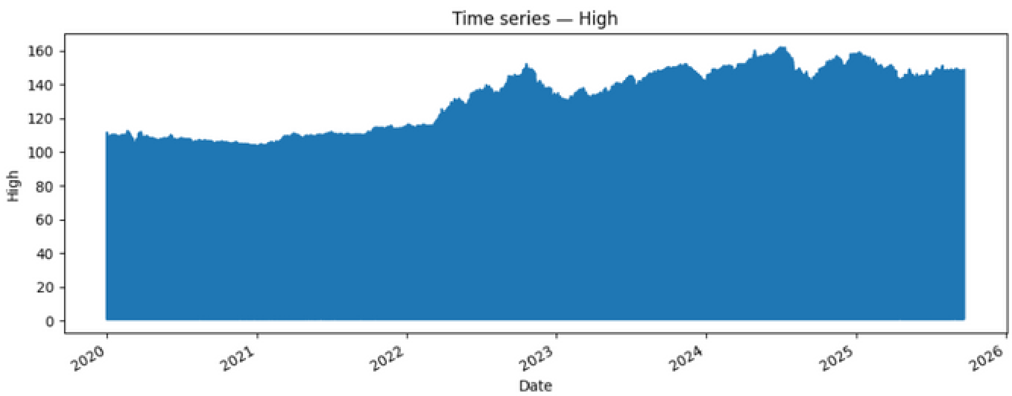
- 100% of the observations are valid,
- No missing values (nulls) are present,
- No errors or inconsistent values were detected across the

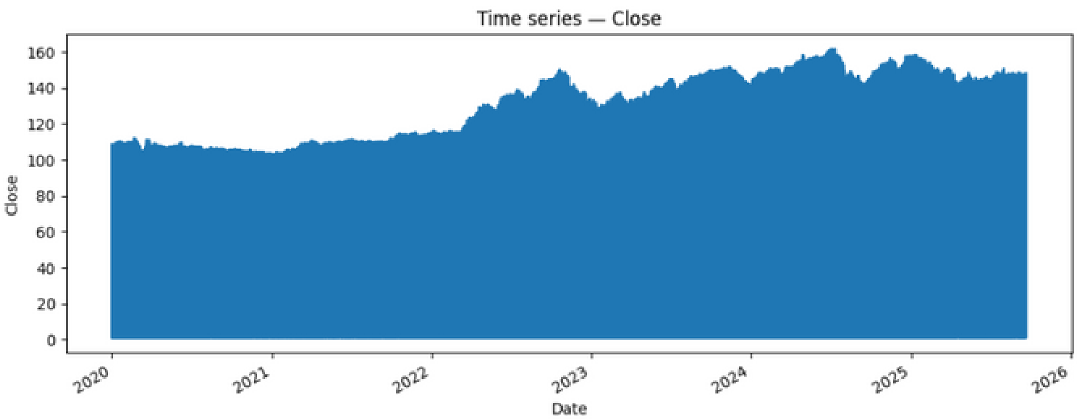
This verification is crucial in time series analysis, as missing values may bias parameter estimation, distort autocorrelation structures, and negatively impact the performance of ARIMA models.

3. Methodology

3.1 Descriptive Statistics







- High instability from 2023 to 2026.
- Clear quarterly seasonality with peaks near quarter end.
- stabilization from 2020 to the middle of 2022.

3.1 Augmented Dickey Fuller Test

```
Test statistic : -1.588247
p-value        : 0.489495
# lags used    : 0
# obs          : 1490
 critical 1% : -3.434746
 critical 5% : -2.863482
 critical 10% : -2.567804
```

- Test statistic : -1.588 :
  less negative than all critical values
- p-value : 0.489 :
  0.489 > 0.05 => fail to reject the hypothesis 0

=> The Augmented Dickey Fuller test fails to rejects the null hypothesis of a unit root at the 5% significance level. Therefore, the time series is non stationary.

The series is not stationary so we should do :

- A. Differencing :

  First-order difference:

  $y_t' = y_t - y_{t-1}$ (each new value is the change from the previous time point.)

  Note : (if first order difference fail we move to a second differencing)

- B. Transformation :

  Log transform: (Reduces variance)

  $y_t' = \ln(y_t)$

1 st order differencing results :

```
Test statistic : -37.796258
p-value        : 0.000000
# lags used    : 0
# obs          : 1489
 critical 1% : -3.434749
 critical 5% : -2.863483
 critical 10% : -2.567804
```

- Test statistic : -37.796 :

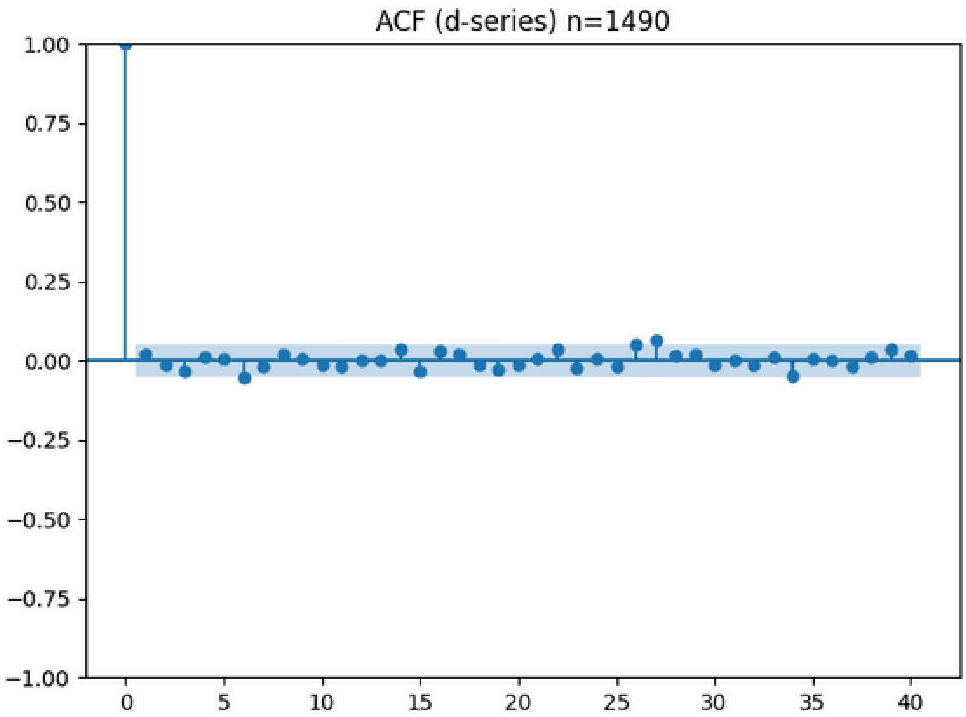
  more negative than all critical values

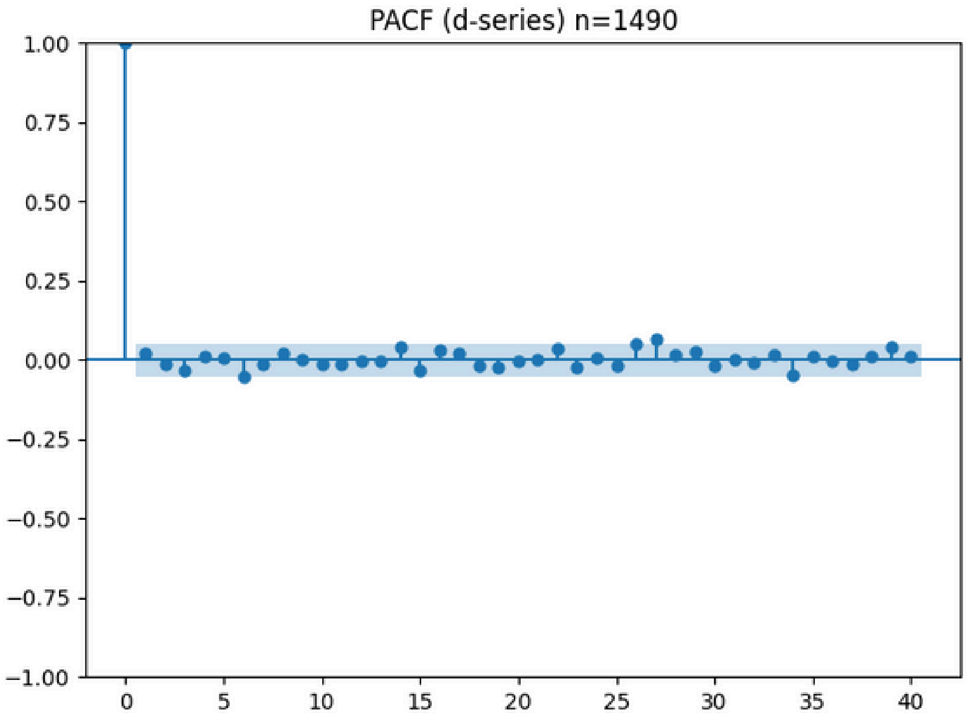
- p-value : 0.000 :

  $0.000 < 0.05$ => reject the hypothesis 0

=> The Augmented Dickey Fuller test succeeded to reject the null hypothesis of a unit root at the 5% significance level. Therefore, the time series is stationary and d=1 is appropriate.

3.2 Identify ARIMA orders (p,d,q)





- 1 at lag 0 (always)
- close to zero for all lags > 0

-The differenced series behaves like white noise (no significant autocorrelation).
-Therefore we choose ARIMA(0,1,0) (random walk) (because there's no significant spikes)

First Choice :

- ARIMA(0,1,0)

Second Choice :

- ARIMA(0,1,1)

Third Choice :

- ARIMA(1,1,0)

3.3 ARIMA Model Selection:

```
Fitted ARIMA(0, 1, 0) — AIC: -11421.085, BIC: -11415.779
```

```
Fitted ARIMA(0, 1, 1) — AIC: -11419.678, BIC: -11409.065
```

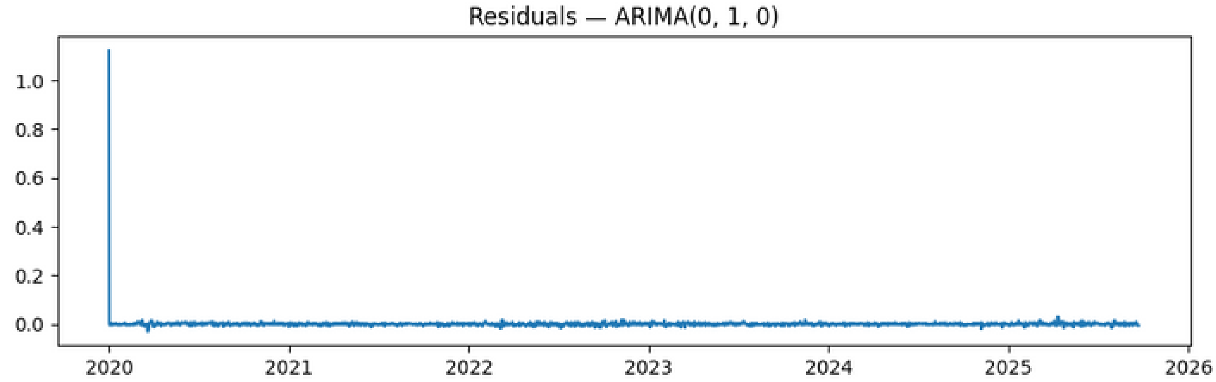
```
Fitted ARIMA(1, 1, 0) — AIC: -11419.665, BIC: -11409.052
```

ARIMA(0,1,0) has the lowest AIC/BIC
=> selected as the best model, which supports our initial analysis
(AR and MA coefficients in the alternatives are extremely close to 0 (≈ 0.019), so adding AR(1) or MA(1) gives almost no extra explanatory)

3.4 Ljung–Box Evaluation Test :

```
Ljung-Box results for residuals (lags 10 and 20):
     lb_stat  lb_pvalue
10  0.124742        1.0
20  0.197018        1.0
```

p-values = 1.0 → fail to reject the null of no autocorrelation at both lags.
=> residuals show no remaining autocorrelation so the model captures the serial dependence perfectly.



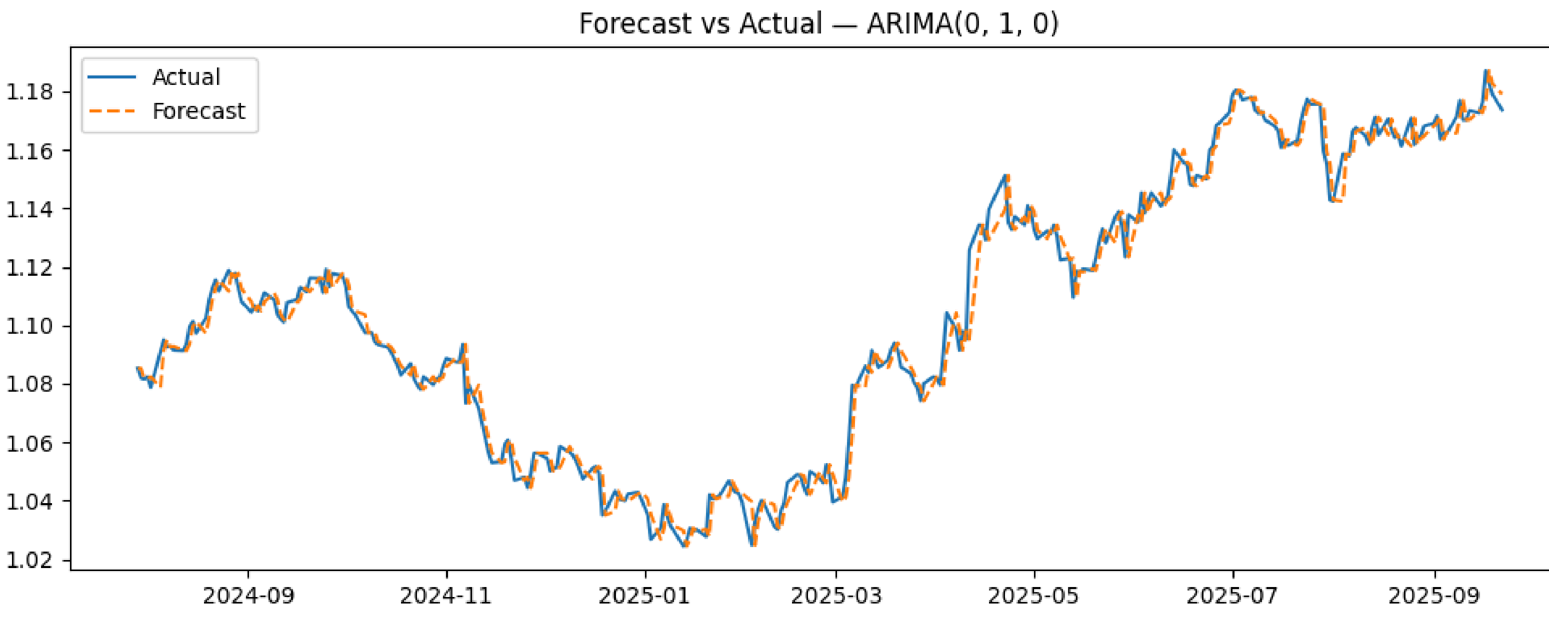
The residuals plot confirms our previous analysis
=> White noise

3.4 Forecasting Results :

```
RMSE = 0.00567853, MAE = 0.00418164
```

- The mean value of the series is 1.1 approximately, so relative RMSE is 4.55% and relative MAE is 3.64% roughly.

=> the model captures the level well.



4.Discussion

- These findings indicate that the dynamics of the series are well described by a random walk (ARIMA(0,1,0)). Original evaluation produced RMSE = 0.05 and MAE = 0.04, which correspond to relative errors of approximately 4.55% (RMSE/mean) and 3.64% (MAE/mean) given a mean level ≈ 1.1 these are small errors in the context of daily FX prices and indicate good one step forecasting performance for the mean.

4.1 Economic Meaning :

- Efforts should pivot from predicting levels to modelling other quantities that are more predictable (volatility) or incorporating exogenous information.
- FX returns commonly display volatility clustering, so we should consider GARCH family models for conditional variance.
- Markov switching to capture different market regimes.

5.Conclusion

- The project applied textbook-standard methods (ADF testing with automatic lag selection, ACF/PACF inspection, ARIMA estimation, Ljung–Box residual checks, and rolling origin evaluation). All evidence shows the series is I(1) and that the differenced series is effectively white noise, consequently ARIMA(0,1,0) (random walk) is the most parsimonious and empirically supported model for the mean. Forecast performance (RMSE = 0.05, MAE = 0.04; relative RMSE ≈ 4.55%) indicates reliable short term point forecasts at the daily frequency, diagnostics confirm no remaining linear autocorrelation in residuals.