

Exchange Rate Forecasting in Algeria (DZD/USD): From the Need for Fourier Stationarity Tests to a Comparison of Traditional and Machine Learning Approaches

Wassila SEMANE*

Faculty of Economic Sciences, Commercial Sciences and
Management Sciences, University of Algiers 3, Algeria
semane.wassila@univ-alger3.dz

Received: 01/05/2026; Accepted: 01/06/2026; Published: 30/06/2026

Abstract:

This article examines the issue of exchange rate forecasting in Algeria in two complementary stages. First, we demonstrate the need to use stationarity tests robust to smooth structural breaks, namely, the Fourier ADF and Fourier KPSS tests—to correctly characterize the dynamics of the series. Second, we compare the predictive performance of classical approaches (SARIMA, exponential smoothing) with that of a machine learning model (Gradient Boosting). The empirical results, based on monthly data from January 1994 to December 2025, confirm that the series is first-order integrated and that the Gradient Boosting model outperforms traditional methods in terms of forecast error (RMSE) and forecast accuracy. This dual approach highlights the importance of flexible modeling both in the analysis of stationarity and in the forecasting of financial series subject to structural changes.

Keywords: exchange rates; Fourier tests; stationarity; structural breaks; forecasting; ARIMA; exponential smoothing; Gradient Boosting.

1.Introduction:

The exchange rate is a fundamental macroeconomic indicator that

* Corresponding author

plays a central role in external competitiveness, the conduct of monetary policy, the attractiveness of international investment, and financial stability. In the case of Algeria, whose economy is heavily dependent on hydrocarbon exports denominated in foreign currencies, particularly the U.S. dollar, the analysis and forecasting of the Algerian dinar exchange rate are of major strategic importance. Since the abandonment of the fixed exchange rate system and the gradual adoption of floating exchange rate regimes beginning in 1973, economies have faced increased exchange rate volatility, making their evolution more difficult to anticipate. Exchange rate forecasting remains a complex exercise due to the multitude of economic and financial factors that influence it, such as the balance of payments, foreign exchange reserves, interest rates, purchasing power, and international capital flows. Faced with this complexity, economic agents seek to anticipate exchange rate movements in order to hedge against exchange rate risk or seize opportunities for short-, medium-, and long-term gains.

In this context, emerging economies such as Algeria face additional challenges: exchange rate policies, recurring oil shocks, and successive economic reforms introduce nonlinearities and structural breaks that affect the dynamics of financial time series. These characteristics make classical linear models, such as ARIMA models or exponential smoothing, potentially limited in their ability to capture the full range of underlying phenomena. In contrast, machine learning methods, thanks to their flexibility and ability to model complex relationships, offer promising prospects for improving forecast accuracy

2- Objective of the study:

The objective of this article is twofold. First, we demonstrate the importance of using stationarity tests that are robust to smooth structural breaks, namely the Fourier ADF and Fourier KPSS tests, to accurately characterize the dynamics of the exchange rate in Algeria.

Second, we compare the predictive performance of three forecasting approaches:

1. The SARIMA model (Box-Jenkins),
2. Exponential smoothing (Holt-Winters),
3. Gradient Boosting (machine learning).

The study is based on monthly exchange rate data for Algeria covering the period from January 1994 to December 2025, sourced from the Bank for International Settlements (BIS). This dual approach—rigorous analysis of stationarity followed by a comparison of forecasting models—provides methodological and empirical insights into the most suitable tools for forecasting financial time series in a context of structural transformations

3- Research Question:

To what extent does the integration of robust stationarity tests for smooth structural breaks (Fourier ADF and Fourier KPSS) and machine learning methods (Gradient Boosting) improve exchange rate forecasting in Algeria compared to traditional approaches (SARIMA and exponential smoothing)?

4- Data and Preliminary Analysis

4.1. Description of the Data

The data series under study covers 384 monthly observations (January 1994 – December 2025). Figure 1 shows the evolution of the exchange rate over this period.

Figure 1: Exchange Rate Trends in Algeria (1994–2025)



Source: Compiled by the author using data from the Bank for International Settlements (BIS).

The chart shows a sharp depreciation in the early 1990s, followed by a period of relative stability between 1995 and 2000, then an appreciation through 2006, and a period of increased volatility after 2008, with marked fluctuations since 2020.

Thus, fluctuations are observed around a downward trend, suggesting a cyclical or seasonal component.

4.2. Stationarity Tests with Structural Breaks : Prior to any modeling, we tested the stationarity of the series using the Fourier ADF and Fourier KPSS⁽¹⁾ tests. These tests allow for the consideration of smooth structural breaks without specifying the break dates a priori.

4.2.1 Fourier ADF Test:

The Fourier ADF test extends the classical ADF test by incorporating sine and cosine trigonometric terms to capture smooth structural changes: $\sin(2\pi kt/T)$ and $\cos(2\pi kt/T)$, where k represents the optimal Fourier frequency (selected by minimizing the sum of the squares of the residuals) and T is the sample size.

➤ Test models :

$$y_t = \gamma_1 \sin\left(\frac{2\pi kt}{T}\right) + \gamma_2 \cos\left(\frac{2\pi kt}{T}\right) + \varepsilon_t$$

$$y_t = \alpha + \gamma_1 \sin\left(\frac{2\pi kt}{T}\right) + \gamma_2 \cos\left(\frac{2\pi kt}{T}\right) + \varepsilon_t$$

$$y_t = \alpha + \beta t + \gamma_1 \sin\left(\frac{2\pi kt}{T}\right) + \gamma_2 \cos\left(\frac{2\pi kt}{T}\right) + \varepsilon_t;$$

➤ Test hypotheses:

- H_0 (null hypothesis): y_t is a non-stationary series (presence of a unit root). Under the null hypothesis, the deterministic component is

(1) Becker, Enders & Lee, A Stationarity Test in the Presence of an Unknown Number of Smooth Breaks. *Journal of Time Series Analysis*, 27(3), pp. 381–409, 2006.

constant (or contains no Fourier terms) and the errors are $I(1)$.

- H_1 (alternative hypothesis): y_t is stationary around a deterministic component represented by a Fourier function. The series therefore returns to a mean that varies smoothly over time.

This specification allows for modeling breaks of unknown form without imposing break dates, which significantly improves the test's performance in the presence of nonlinearities.

4.2.2 Fourier KPSS Test:

The KPSS test with Fourier components (Fourier KPSS) is designed to improve upon the standard KPSS test by allowing for smooth, unknown structural changes in the deterministic component of the time series.

By incorporating low-frequency trigonometric terms, the test approximates a wide range of possible structural breaks without imposing discrete break points. This eliminates the need to specify break dates in advance, which is often a significant limitation in traditional structural break models.

Consequently, the KPSS test with Fourier components exhibits superior power and robustness when testing for stationarity in the presence of nonlinear dynamics and gradual regime shifts.

The three models of the standard KPSS test (Kwiatkowski-Phillips-Schmidt-Shin, 1992) are as follows:

$$y_t = \alpha + r_t + \beta * t + \varepsilon_t$$

$$y_t = \alpha + \gamma_1 \sin\left(\frac{2\pi kt}{T}\right) + \gamma_2 \cos\left(\frac{2\pi kt}{T}\right) + \varepsilon_t$$

$$y_t = \alpha + \beta t + \gamma_1 \sin\left(\frac{2\pi kt}{T}\right) + \gamma_2 \cos\left(\frac{2\pi kt}{T}\right) + \varepsilon_t$$

➤ the hypotheses:

- Null hypothesis (H_0): y_t is stationary around its Fourier components.

- Alternative hypothesis (H_1): yt has a unit root (non-stationary).

This formulation corresponds to the inversion of the hypotheses relative to the standard KPSS test. Indeed, in the Fourier KPSS test (unlike the classical KPSS where H_0 is stationarity), we retain H_0 (stationarity), but this is now defined around a trend modeled by Fourier components.

➤ Test Statistics:

$$\tau = \frac{1}{T^2} \mathbf{1}' * \sum \frac{s_t^2}{\sigma_{lr}^2}$$

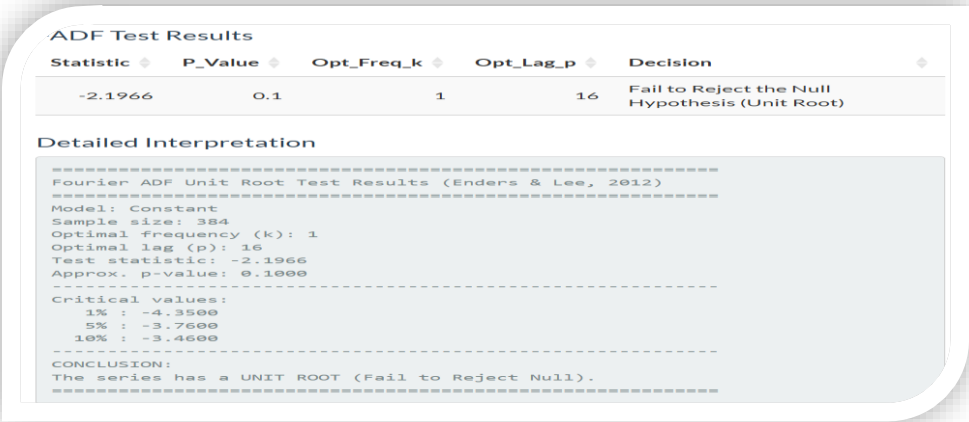
Where:

$S_t = \sum_{j=1}^t \hat{e}_j$ corresponds to the partial sum of the residuals

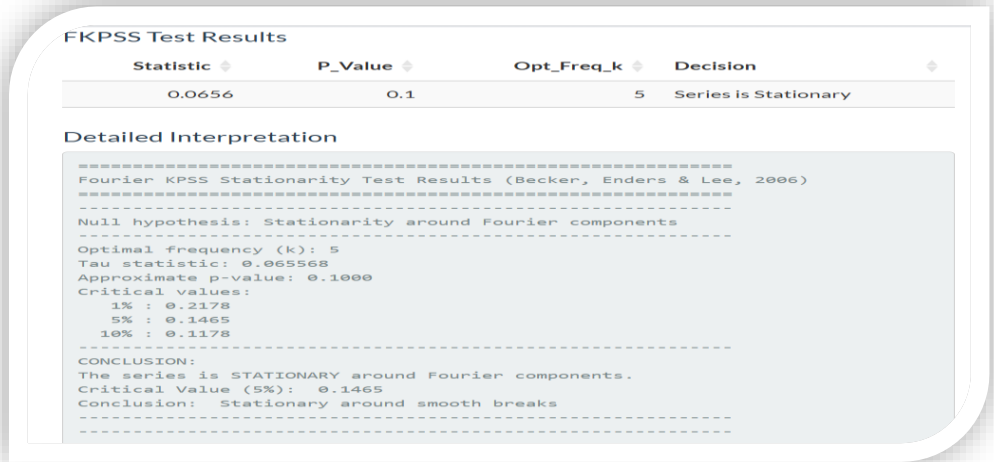
σ_{lr}^2 is the long-term variance.

4.2.3 Test results:

➤ On the original series:



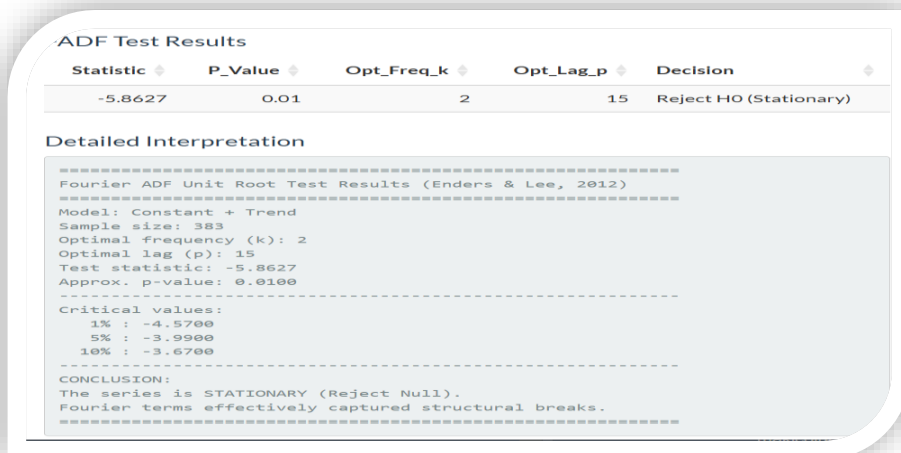
Source: Generated by the author using Python



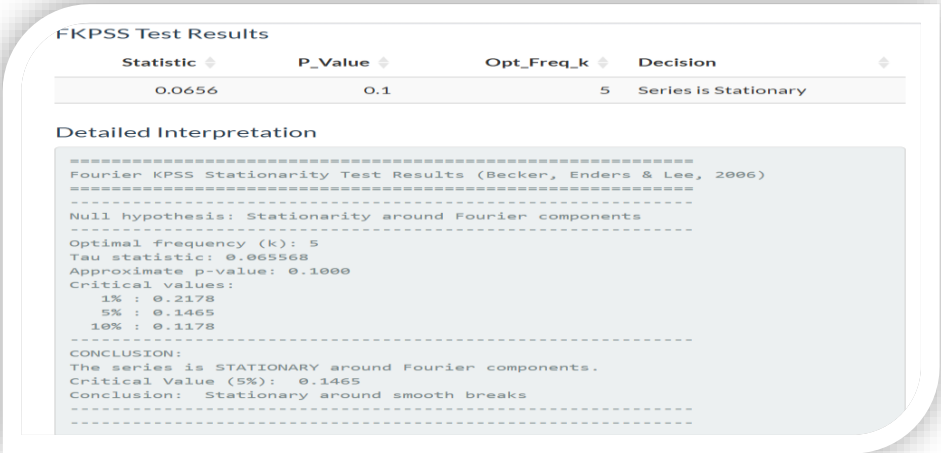
Source: Generated by the author using Python

➤ **For the first-order differential series:**

The results indicate that it becomes stationary after the first differentiation.

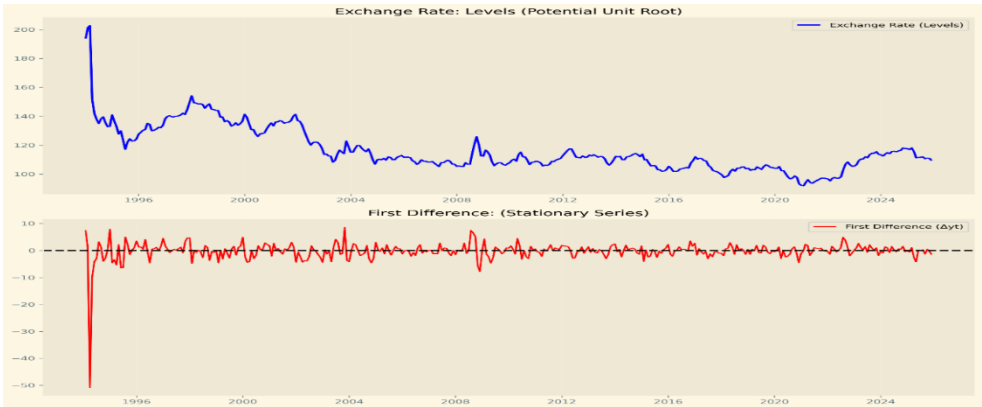


Source: Generated by the author using Python



Source: Generated by the author using Python

A joint analysis of the results of the ADF and KPSS tests confirms that the series is non-stationary in level but becomes stationary after differentiation, indicating that it is integrated of order one. The use of Fourier approaches improves the robustness of the tests in the presence of smooth structural breaks. The Following graph clearly shows stationarity after the first differentiation.



Source: Graph generated by the author with Python

5. Literature Review on Forecasting Methodology:

5.1. Traditional Approaches

5.1.1. The Box-Jenkins Methodology⁽²⁾

a) Definition:

The Box-Jenkins methodology is a statistical approach to short-term forecasting developed by George Box and Gwilym Jenkins in 1976. Based on autoregressive and moving average models, it relies on advanced mathematical and statistical principles. Its use is primarily geared toward short-term forecasting, as the accuracy of results tends to decrease as the forecast horizon lengthens due to the accumulation of errors and the complexity of temporal dynamics. The development of statistical software has enabled a broader and more effective empirical application of this methodology.

b) Conditions for applying the Box-Jenkins methodology:

The Box-Jenkins methodology is based on a set of essential conditions that must be met to ensure the validity and reliability of the forecast results. First, each time series is analyzed independently. Second, the series under study must be stationary, a prerequisite for the correct estimation of the models. Furthermore, the data must not contain any missing values, to avoid any distortion in the analysis. Finally, the methodology assumes that the dynamics of the series can be represented by a linear process.

c) Box-Jenkins models:

Models derived from the Box-Jenkins methodology fall into two broad categories. The first category includes non-seasonal models, namely autoregressive (AR), moving average (MA), mixed (ARMA), and integrated (ARIMA) models. The second category concerns seasonal models, such as SAR, SMA, SARMA, and SARIMA models, which explicitly incorporate the seasonal component of time series.

Autoregressive (AR(p)) and moving average (MA(q)) models form the foundation of time series analysis.

(2) Bourbonnais, R. *Économétrie*, 8e édition, Dunod, Paris, pp. 235-270 , 2010.

An AR(p) model explains the current value of a series as a linear combination of its own lagged values, while an MA(q) model describes the current value of a time series as a linear combination of present and past random errors over a finite number of periods (Box et al., 2015).

Formally, an AR(p) model is written as:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3} + \dots + \theta_p y_{t-p} + \varepsilon_t$$

whereas the MA(q) model is given by :

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

where ε_t is white noise.

MA(q) models are characterized by an autocorrelation function that vanishes after a lag of q, whereas AR(p) models exhibit a gradual decline in autocorrelation

The combination of autoregressive and moving average components leads to ARMA models, and their extension to non-stationary series gives rise to ARIMA models, which are widely used for the identification and forecasting of economic and financial time series⁽³⁾

The SARIMA (Seasonal Autoregressive Integrated Moving Average) model is an extension of the ARIMA model by Box and Jenkins (1976)⁽⁴⁾ that allows for the seasonal component present in monthly time series. It is characterized by three main components: an autoregressive (AR) component, a moving average (MA) component, and their seasonal equivalents (SAR, SMA), all associated with an order of differentiation (d) that ensures the stationarity of the series.

d) Characteristics of Box-Jenkins Models:

The Box-Jenkins approach consists of four steps:

(3) Hamilton, J. D. Time Series Analysis, Princeton University Press, pp. 100–120, 1994

(4) Box, G. E. P., & Jenkins, G. M. Time Series Analysis: Forecasting and Control, Holden-Day, pp. 300–315, 1976

identification, visual inspection (ACF/PACF), estimation (maximum likelihood or least squares methods), diagnosis (the model coefficients must be significantly different from 0; normality tests for the residuals, whiteness tests), and forecasting.

The choice of orders p and q is based on the analysis of autocorrelation and partial autocorrelation functions as well as on information criteria (AIC, BIC). A key advantage of Box-Jenkins models is their ability to model different components of time series dynamics in a parametric form, but this strength requires a rigorous examination of the residuals (normality tests, Ljung-Box) and a check of parameter stability over time.

For example, empirical studies on exchange rates show that well-specified ARIMA models can provide accurate forecasts over very short horizons, provided that the time series is properly transformed and that structural breaks are taken into account

5.1.2 Holt-Winters Exponential Smoothing⁽⁵⁾:

Developed by Holt and Winters in 1960, this method is the preferred choice for time series exhibiting both a trend and seasonality. It is based on the idea that past observations can be used to predict future values by assigning an exponentially decreasing weight to the oldest observations.

It involves estimating three components: the level, the trend, and the seasonality. There are three versions, two of which are the main ones: one suited for series with a multiplicative decomposition, the other for series with an additive decomposition.

This is a widely used forecasting technique for time series that exhibit both a trend and seasonality; as such, this flexible and intuitive method is widely used in the fields of inventory management, sales, economics, and budget planning.

(5) Mélard, G. Méthodes de prévision à court terme, Editions Ellipses, Paris, p. 159.,2007

➤ **Holt-Winters Method (multiplicative version)**

In this case, the series can be written as:

$$X_t = (a_t + b_t) * S_t + \varepsilon_t$$

Three smoothing filters are applied, with smoothing coefficients α , β and γ ranging from 0 to 1:

➤ **Level Smoothing a_t :**

$$a_t = \alpha \left(\frac{x_t}{S_{t-p}} \right) + (1 - \alpha)(a_{t-1} + b_{t-1})$$

➤ **Trend Smoothing b_t :**

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$$

➤ **Seasonal Smoothing S_t :**

$$S_t = \gamma \left(\frac{x_t}{a_t} \right) + (1 - \gamma)S_{t-p}$$

➤ **Forecasts:**

For a forecast over a horizon of h (where $1 \leq h \leq p$)

$$\hat{X}_{t+h} = (a_t + hb_t) \times S_{t-p+h}$$

For ($p+1 \leq h \leq 2p$):

$$\hat{X}_{t+h} = (a_t + hb_t) \times S_{t-2p+h}$$

The initial values are calculated as follows:

- Level: $a_p = \bar{x}$ (first-year average)

- Trend: $b_p = 0$

- Seasonality for $t = 1, \dots, p$:

$$S_t = \frac{x_t}{\bar{x}}$$

The seasonal coefficients must satisfy $\sum_{j=1}^p S_j = 1$. If this is not the case, they are adjusted to satisfy this condition.

➤ **Holt-Winters method (additive version):**

In this version, the time series is modeled as follows:

$$X_t = (a_t + b_t \cdot t) + S_t + E_t$$

The smoothing equations are as follows:

➤ **Level Smoothing a_t :**

$$a_t = \alpha(X_t - S_{t-p}) + (1 - \alpha)(a_{t-1} + b_{t-1})$$

➤ **Trend Smoothing b_t :**

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta) \cdot b_{t-1}$$

➤ **Seasonal Smoothing S_t**

$$S_t = \gamma(X_t - a_t) + (1 - \gamma)S_{t-p}$$

➤ **Forecasts:**

For a forecast over a horizon of h (where $1 \leq h \leq p$):

$$\hat{X}_{t+h} = (a_t + h \cdot b_t) + S_{t-p+h}$$

For ($p+1 \leq h \leq 2p$) :

$$\hat{X}_{t+h} = (a_t + h \cdot b_t) + S_{t-2p+h}$$

The initial values are calculated as follows:

- Level: $a_p = \bar{x}$ (first-year average)

- Trend: The trend is estimated based on the difference between the averages of the first two cycles:

$$b_{p=} \frac{\bar{x}_2 - \bar{x}_1}{p}$$

where $\bar{x}_2$ is the average for the second year (period $p+1$ to $2p$)

- Seasonality for $t = 1, \dots, p$:

We calculate the difference between each observation in the first year and the average for that year

The seasonal coefficients must satisfy

$$S_t = x_t - \bar{x}_1 \quad \text{pour } t=1,2,\dots,P$$

If this is not the case, they are adjusted to meet this condition.

5.2 Machine Learning Approach:

Ensemble learning involves combining multiple predictive models to achieve a performance that exceeds that of any single model. Three

major contributions established the theoretical foundations of this field between 1996 and 2001: Bagging proposed by Leo Breiman, Random Forests also developed by Breiman, and Gradient Boosting formalized by Jerome Friedman.

Although these methods share the common goal of improving predictive accuracy by combining models, their architectures and theoretical foundations differ profoundly. Bagging reduces variance by aggregating unstable models trained on bootstrap samples. Random Forests extend this principle to decision trees by introducing feature randomization, which decorrelates the models. Gradient Boosting, on the other hand, sequentially constructs weak models by following the negative gradient of the loss function, thereby targeting the reduction of bias.

5.2.1 Common Mathematical Foundations:

➤ The Prediction Problem

Let be a training set :

$$\mathcal{L} = \{(\mathbf{y}_n, \mathbf{x}_n), n = 1, \dots, N\} \text{ où } \mathbf{x} \in \mathcal{X}$$

Let $\mathbf{x}$ be the predictor vector and $y \in Y$ the response variable. The objective is to construct a predictor $\phi(\mathbf{x}, \mathcal{L})$ that minimizes the expected prediction error.

➤ Bias-Variance Decomposition

For a regression problem with a quadratic loss function, the mean squared error (MSE) decomposes into:

$$\begin{aligned} \mathbb{E}_{\mathcal{L}}[(\mathbf{y} - \phi(\mathbf{x}, \mathcal{L}))^2] &= \underbrace{(\mathbf{y} - \mathbb{E}_{\mathcal{L}}[\phi(\mathbf{x}, \mathcal{L})])^2}_{\text{Bias}^2} \\ &= \underbrace{(\mathbf{y} - \mathbb{E}_{\mathcal{L}}[\phi(\mathbf{x}, \mathcal{L})])^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}_{\mathcal{L}}[(\phi(\mathbf{x}, \mathcal{L}) - \mathbb{E}_{\mathcal{L}}[\phi(\mathbf{x}, \mathcal{L})])^2]}_{\text{variance}} \end{aligned}$$

This decomposition structures the improvement strategies: reduce bias or reduce variance

5.2.2 Bagging Predictors⁽⁶⁾

➤ Definition and Mechanism:

Bagging (Bootstrap Aggregating) is based on a simple yet powerful principle: rather than building a single predictor on the training set L , we generate B Bootstrap replicas $L^{(b)}$ via with-replacement sampling, train a predictor $\phi(x, L^{(b)})$ on each one, and aggregate the results.

For regression:

$$\phi_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \phi(x, L^{(b)})$$

For classification:

$$\phi_{bag}(x) = \arg \max_{y \in Y} \sum_{b=1}^B \mathbf{1}[\phi(x, L^{(b)}) = y]$$

➤ Theoretical Foundation:

Breiman demonstrates that aggregation systematically reduces the mean squared error. By defining the ideal aggregated predictor :

$$\phi_A(x) = \mathbb{E}_L[\phi(x, L)]$$

Jensen's inequality leads to :

$$\mathbb{E}_L[(y - \phi(x, L))^2] \geq \underbrace{(y - \phi_A(x))^2}$$

The improvement achieved through aggregation is exactly equal to the variance of the individual predictors:

$$\text{Gain} = \mathbb{E}_L[(\phi(x, L))^2] - (\mathbb{E}_L[\phi(x, L)])^2 = \text{Var}_L(\phi(x, L))$$

➤ Conditions for Effectiveness

Bagging is particularly effective when the base predictor is unstable, that is, sensitive to small perturbations in the training set. Decision trees and neural networks exhibit this property, whereas

(6) Breiman, L. Bagging predictors. *Machine Learning*, 24(2), pp. 123–140, 1996

stable models (linear regression, k-nearest neighbors) benefit little from Bagging.

5.2.3. Random Forests⁽⁷⁾

➤ Conceptual Architecture:

Random Forests extend Bagging to decision trees by adding a second source of stochasticity. A Random Forest is a collection of trees

$\{h(x, \theta_k), k = 1, \dots, K\}$, où θ_k Is an i.i.d. random vector parameterizing:

- The Bootstrap sample used to train the tree;
- The random selection of a subset of features at each node.

➤ The Margin Function:

For a classification problem, Breiman defines the margin function:

$$mg(X, Y) = \frac{1}{K} \sum_{k=1}^K \mathbf{1}[h_k(X) = Y] - \max_{j \neq Y} \frac{1}{K} \sum_{k=1}^K \mathbf{1}[h_k(X) = j]$$

The margin measures the confidence in the prediction: a positive value indicates that the majority vote is in favor of the correct class.

➤ Convergence of the Generalization Error :

Breiman demonstrates that as the number of trees increases, the generalization error converges almost surely to:

$$PE^* = \mathbb{P}_{X,Y}(mg(X, Y) < 0)$$

This fundamental property establishes that Random Forests do not overfit as more trees are added - the generalization error reaches an asymptotic limit.

➤ Strength-Correlation Trade-off :

(7) Breiman, L. Random forests. Machine Learning, 45(1), pp. 5–32, 2001

The crucial innovation of Random Forests lies in the explicit control of the correlation between trees. The randomization of features reduces inter-tree correlation, thereby allowing for a low generalization error even when the number of trees is large.

5.2.4 Gradient Boosting⁽⁸⁾

➤ General Principle:

Unlike parallel methods such as Bagging and Random Forests, Gradient Boosting takes a sequential approach. The final model is constructed as a weighted sum of weak models:

$$F_M(x) = \sum_{m=1}^M p_m h(x; a_m)$$

where $h(x; a_m)$ is a weak model (typically a shallow tree) and p_m is its coefficient.

➤ Functional Gradient Descent :

Friedman's major theoretical innovation is the reformulation of boosting as a functional optimization problem solved by gradient descent. At each iteration m , the pseudo-residual is calculated:

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F=F_{m-1}}$$

The weak h_m model is then trained to predict these pseudo-residues, and the coefficient p_m is determined by a linear search:

$$p_m = \arg \min \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + p h_m(x_i))$$

The update is then performed by:

$$F_m(x) = F_{m-1}(x) + v p_m h_m(x)$$

where $v \in (0,1]$ is the learning rate that regularizes the learning

(8) Friedman, J. H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), pp. 1189–1232, 2001

process

➤ **Loss Functions:**

The Gradient Boosting framework applies to any differentiable and convex loss function, making it extremely flexible:

Problem	Loss Function	Pseudo-Residual
Regression	$\frac{1}{2}(Y_i - F)^2$	$Y_i - F(x_i)$
Binary Classification	$-Y \log(p) - (1-y) \log(1-p)$	$Y_i - p(x_i)$
Robust Regression	$ Y_i - F(x_i) $	$\text{sign}(Y_i - F(x_i))$

Source: Author’s compilation based on Friedman (2001)

➤ **Overfitting Control:**

Unlike Random Forests, Gradient Boosting can overfit if the number of iterations is too high. The main regularization mechanisms are:

- The learning rate ν (low values slow down learning and improve generalization);
- Tree depth (generally limited to 3–8 levels);
- Subsampling (stochastic gradient boosting), which introduces stochasticity similar to Bagging.

5. Results and Interpretation:

The evaluation of the forecasting models is based on a systematic comparison of six distinct methodological approaches, covering both classical exponential smoothing and ARIMA modeling methods, as well as ensemble learning methods (Bagging, Random Forests, AdaBoost, Gradient Boosting)

To assess the generalization ability of the estimated models, the total sample is partitioned according to a standard procedure in supervised learning: 80% of the observations are allocated to the training set and 20% to the testing set. This partition allows for the

measurement of ex ante predictive performance on data not used during the estimation phase, thereby avoiding the optimism bias associated with an evaluation on the training set.

Three metrics are used to evaluate the accuracy of the forecasts:

- RMSE (*Root Mean Squared Error*):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- MAE (*Mean Absolute Error*):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- MAPE (*Mean Absolute Percentage Error*) :

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

These metrics provide a comprehensive view of forecast error, with RMSE penalizing large errors more heavily, while MAPE provides a relative measure that is independent of the variable's scale

5.1 Selection of the ARIMA/SARIMA Model :

The search for the best ARIMA model is conducted using a stepwise search procedure that minimizes the Akaike Information Criterion (AIC). This approach systematically explores a set of combinations of autoregressive orders (p), integration orders (d), and moving averages (q), as well as their seasonal equivalents (P, D, Q) for a monthly periodicity (s = 12).

The main results of this search are as follows:

Model	AIC	Constant
ARIMA(2,1,0)(1,0,1)[12]	1656.636	No
ARIMA(2,1,0)(1,0,0)[12]	1657.030	No
ARIMA(2,1,0)(2,0,1)[12]	1657.568	No

ARIMA(2,1,1)(1,0,1)[12]	1658.615	No
ARIMA(3,1,1)(1,0,1)[12]	1658.494	No

The model selected is therefore a **SARIMA(2,1,0)(1,0,1)[12]** model without a constant term. This specification is interpreted as follows:

- **Non-seasonal component** : an integration order of $d=1$ (the series is differentiated once to become stationary), two autoregressive terms ($p=2$), and no non-seasonal moving average ($q=0$).
- **Seasonal component**: no seasonal integration order ($d=0$), one seasonal autoregressive term ($p=1$), and one seasonal moving average term ($q=1$).

The absence of a constant term in the final model is consistent with first-order differentiation: an integrated series generally does not exhibit systematic drift after differentiation.

5.2 Comparative Analysis of Predictive Performance :

Model	RMSE	MAE	MAPE (%)
Gradient Boosting	2.2040	1.3159	1.35%
Bagging	2.5225	1.6189	1.66%
Random Forest	2.6327	1.8195	1.84%
AdaBoost	4.2323	3.0416	3.08%
Exponential Smoothing	8.1310	7.2181	6.88%
Auto-ARIMA	8.5834	7.6425	7.13%

The results reveal a clear hierarchy among the different classes of models. Ensemble learning methods (Gradient Boosting, Bagging, Random Forests, AdaBoost) significantly outperform traditional methods (exponential smoothing and ARIMA). This finding confirms the value of ensemble approaches for modeling complex time series, particularly when exogenous variables are available.

Gradient Boosting stands out as the best-performing model across all three metrics, with an RMSE of 2.2040, an MAE of 1.3159, and a MAPE of 1.35%.

This superiority can be explained by its ability to model nonlinear relationships and its sequential optimization mechanism, which focuses the learning effort on the observations that are most difficult to predict.

6.3 Forecasting:

The selected Gradient Boosting model is used to generate point forecasts for the next five months. The results are presented in the table below:

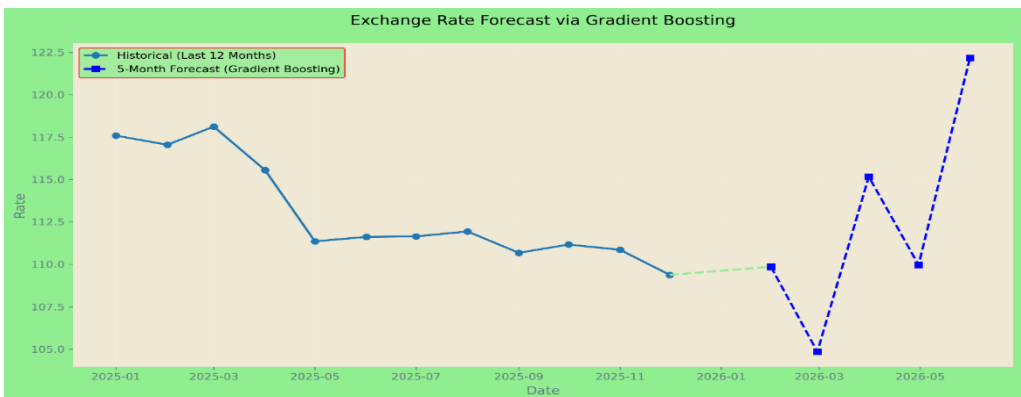
Month 1: 109.8597

Month 2: 104.8481

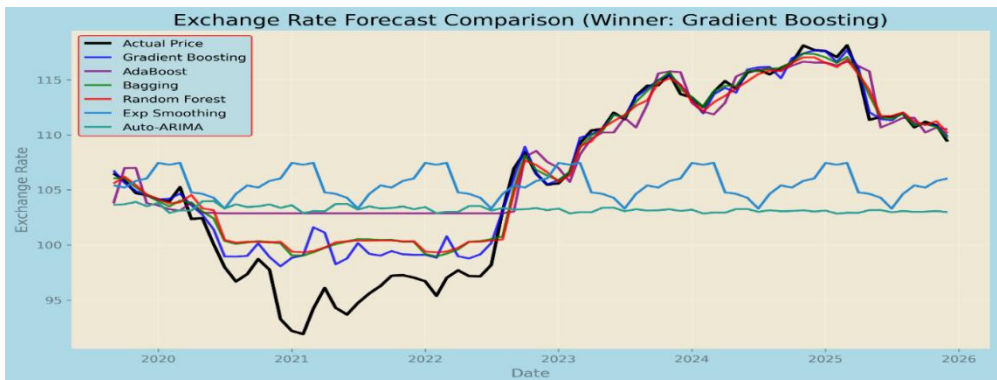
Month 3: 115.1705

Month 4: 109.9665

Month 5: 122.165



Source: Graph generated by the author with Python



Source: Graph generated by the author with Python

7. Conclusion:

This article aimed to provide methodological and empirical insights into exchange rate forecasting in Algeria by combining two complementary steps. The first step involved accurately characterizing the dynamics of the time series using stationarity tests robust to smooth structural breaks, namely the Fourier ADF and Fourier KPSS tests. The second step focused on comparing the predictive performance of classical approaches (SARIMA and Holt-Winters exponential smoothing) with that of a machine learning model, Gradient Boosting, representative of ensemble methods.

Analysis of the monthly DZD/USD exchange rate data, covering the period from January 1994 to December 2025, yields three main findings:

- ❖ applying the ADF and KPSS tests to the level series shows that it is non-stationary, while the same tests applied to the first-order differenced series indicate stationarity. This dual procedure validates the first-order integrated nature of the series and underscores the importance of using tests capable of accounting for smooth structural breaks, which are common in financial series from emerging economies subject to exogenous shocks and successive reforms.

- ❖ a comparison of forecasting models reveals a clear hierarchy among the model classes. Machine learning methods (Gradient Boosting, Bagging, Random Forests, AdaBoost) significantly outperform traditional methods (SARIMA and exponential smoothing) across all evaluation metrics (RMSE, MAE, MAPE). This result confirms the value of using nonlinear and multivariate approaches for modeling complex time series, particularly when exogenous variables—such as lagged exchange rates, moving window indicators, or seasonality, are likely to provide additional predictive information.

❖ the Gradient Boosting model stands out as the best performer, with an RMSE of 2.2040, an MAE of 1.3159, and a MAPE of 1.35%. This superiority stems from its ability to model nonlinear relationships and its sequential optimization mechanism, which focuses the learning effort on the observations that are most difficult to predict. The forecasts obtained for the next five months show inter-monthly variability consistent with the seasonal structure of the series, but their interpretation calls for methodological caution in the absence of confidence intervals.

8. Bibliography List

- Afriyie, J. K., Twumasi-Ankrah, S., Gyamfi, K. B., Arthur, D., & Pels, W. A. Evaluating the performance of unit root tests in single time series processes. *Mathematics and Statistics*, 8(6), 2020
- Becker, Enders & Lee, A Stationarity Test in the Presence of an Unknown Number of Smooth Breaks. *Journal of Time Series Analysis*, 27(3), pp. 381–409, 2006.
- Bourbonnais, R. *Économétrie*, 8e édition, Dunod, Paris, pp. 235–270, 2010.
- Box, G. E. P., & Jenkins, G. M. *Time Series Analysis: Forecasting and Control*, Holden-Day, pp. 300–315, 1976
- Breiman, L. Bagging predictors. *Machine Learning*, 24(2), pp. 123–140, 1996
- Breiman, L. Random forests. *Machine Learning*, 45(1), pp. 5–32, 2001
- Friedman, J. H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), pp. 1189–1232, 2001
- Hamilton, J. D. *Time Series Analysis*, Princeton University Press, pp. 100–120, 1994
- Mélard, G. *Méthodes de prévision à court terme*, Editions Ellipses, Paris, p. 159., 20