

Comparative Analysis of ARDL, LSTM, and XGBoost Models For Forecasting The Moroccan Stock Market During The COVID-19 Pandemic

Oukhouya Mohamed Hassan¹, Angour Nora², Aboutabit Nouredine¹, Hafidi Imad¹

¹Laboratory LIPIM, ENSA Khouribga, University Sultan Moulay Slimane, Khouribga, Morocco

²Faculty of Law, Economics and Social Sciences, Salé, Mohammed V University, Rabat, Morocco

E-mail: oukhouya.mhassan@gmail.com, n.angour@um5r.ac.ma, n.aboutabit@usms.ma, i.hafidi@usms.ma

Keywords: ARDL, COVID19, MASI, LSTM, XGBOOST, forecasting

Received: February 22, 2024

This study evaluates and compares the forecasting performances of the ARDL (AutoRegressive Distributed Lag), LSTM (Long Short-Term Memory), and XGBOOST (Extreme Gradient Boosting) models on the MASI (Moroccan All Shares Index). The analysis incorporates daily new COVID-19 cases into the ARDL approach to investigate short-term and long-term relationships with MASI. Cointegration and causality tests are conducted on daily time series data. In terms of accuracy, the ARDL model, especially when including trend and seasonality variables, outperforms LSTM and XGBOOST models. The ARDL model with lags, trend, and seasonality variables achieves the lowest Mean Absolute Percentage Error (MAPE) of 26.7%, with a processing time of 1 second. In comparison, the LSTM and XGBOOST models have MAPE values of 30.5% and 32%, respectively, while requiring significantly longer processing times. These findings suggest that the ARDL model is more efficient and accurate in predicting future values of MASI under pandemic conditions.

Povzetek: Članek primerja modele ARDL, LSTM in XGBoost za napovedovanje maroškega indeksa MASI med COVID-19. Model ARDL presega modele globokega učenja glede natančnosti in učinkovitosti, kar poudarja odpornost ekonometričnih pristopov pri finančnih napovedih.

1 Introduction

Investing in financial markets has long been a focal point for capital holders. While ongoing research and development of new strategies persist, the ever-evolving nature of financial markets necessitates constant adaptation by traders. Consequently, they are increasingly turning to machine learning as a means to enhance and optimize their trading systems. This approach allows them to harness the power of advanced algorithms and data analysis techniques to stay competitive and make more informed investment decisions in the dynamic world of finance.

As financial markets keep changing with more traders and a lot of money involved, it becomes extremely hard to perfectly predict what prices will do in the future. These markets are very complicated, and many things can affect prices, making it almost impossible to guess accurately. That's why traders use complex models and strategies to handle this uncertainty [15]. The various factors that affect financial time series create a situation where they don't stay the same over time, making them non-stationary. This non-stationarity adds complexity to the already challenging tasks of predicting future outcomes and making smart investments in financial markets. Researchers are keenly interested in finding ways to transform these non-stationary time series into more predictable and stable ones, as this can greatly enhance the effectiveness of trading systems [11].

Successful financial forecasting involves combining financial theory, market analysis, diverse data sources, and computational advancements. To build precise models, a comprehensive toolkit is essential, with financial econometrics and machine learning being key components. Machine learning enables better predictions by handling non-linear data, complex variable interactions, and unstructured datasets. However, financial econometrics remains crucial for inferential analysis of economic relationships in finance, and its importance persists alongside machine learning [5]

Financial markets are continuously impacted by events around the world (war, pandemic, natural disaster, etc.). The economic impact of COVID-19 via government actions was examined and it was proved that stocks are negatively affected by social distancing, but positive outcomes are observed for awareness, testing, and income support measures, thereby highlighting the dual economic effects of government responses [2]. As an example of war impact, the Russian-Ukrainian conflict has influenced the interconnections between Russia, European financial markets, and global commodity markets [17].

Time series forecasting is a critical tool in predicting future outcomes in various fields. In the finance industry, forecasting financial market trends is essential for investors, policymakers, and analysts to make informed decisions. The recent COVID-19 pandemic has caused significant disruptions to global financial markets, including the

Moroccan financial market. The impact of the pandemic on the financial market has made forecasting more challenging, and traditional methods may not be enough to capture the complexity of the situation.

Machine learning algorithms have shown promise in providing accurate predictions for various forecasting problems, including time series forecasting. In this article, we aim to explore the effectiveness of machine learning algorithms in forecasting the Moroccan financial market's trends, with a particular focus on the impact of the COVID-19 pandemic.

We will explore the hypothesis that machine learning algorithms, specifically ARDL, LSTM, and XGBoost, can provide more accurate and reliable predictions for the Moroccan financial market than traditional methods. The article aims to contribute to the existing body of knowledge by providing insights into the effectiveness of machine learning algorithms in financial time series forecasting and their ability to capture the effects of unpredictable events such as the COVID-19 pandemic.

2 Related work

Forecasting exchange rates is a critical task in the financial industry, and it has attracted significant attention from researchers in recent years. The high volatility and complexity of the foreign exchange market make it challenging to predict exchange rates accurately. Deep learning techniques, particularly LSTM and XGBoost, as well as the ARDL model, have been widely used to tackle this problem. In this section, we will discuss some works that studied the predictability of Forex based on LSTM, XGBoost, and ARDL.

The article [6] examines the impact of the COVID-19 pandemic on the Saudi Arabian stock market. The authors use an ARDL model to analyze the relationship between COVID-19 cases and the stock market index, taking into account other factors such as oil prices and exchange rates. They use data from January 2015 to July 2020, which includes the period of the pandemic. The results show that COVID-19 has had a significant negative impact on the Saudi stock market, with a decrease in the stock market index following an increase in the number of COVID-19 cases. The authors also find that oil prices and exchange rates have a significant impact on the stock market, but that the effect of COVID-19 is larger. The article provides evidence of the impact of the COVID-19 pandemic on the Saudi Arabian stock market and highlights the importance of considering multiple factors when analyzing the relationship between the stock market and external events. The findings may be useful for investors and policymakers in the region.

This study [14] examines the impact of COVID-19 on daily market returns in affected developed and emerging markets. It finds that an increase in new cases and deaths negatively affects market returns globally. Interestingly,

daily testing has a positive impact. These findings apply to both developed and emerging markets, with the exception that news of new COVID-19 deaths positively affects emerging markets. The study suggests that early proactive measures by governments can protect financial markets and boost investor confidence during future pandemics.

This study [10] investigates the correlation between stock market progress, economic growth, and financial innovation in Bangladesh from 1980 to 2016. To examine the cointegration in the long term, the ARDL bounds test was utilized. Additionally, the Granger Causality test was implemented to identify directional causality between the variables through the error correction mechanism. The ARDL bounds test approach confirms the existence of a long-term relationship between economic growth, stock market progress, and financial innovation. Moreover, the results of the Granger Causality test show a mutual relationship between financial innovation, economic growth, and stock market progress, both in the long and short run. These findings affirm the hypothesis that the growth of market-based financial systems and financial innovation can drive economic growth.

The authors in [13] analyze the connection between oil prices and stock prices in both oil-exporting and oil-importing countries. To do so, they approach the relationship from several angles. Firstly, they examine the possibility of non-linearities in the relationship to determine the unequal response of stock prices in the two categories to positive and negative shifts in oil prices. Secondly, they account for within-group differences by allowing for heterogeneity in the cross sections. Thirdly, they compare the predictability of linear (symmetric) and nonlinear (asymmetric) Panel ARDL models through the Campbell and Thompson (2008) test. The results reveal that the stock prices in both oil-exporting and oil-importing countries respond differently to changes in oil prices, with a stronger response seen in oil-importing countries compared to oil-exporting countries.

The article [12] introduces a new model for forecasting nonlinear time series data. The proposed model combines two existing techniques: EMD (empirical mode decomposition) and NARDL (neural autoregressive distributed lag) modeling. EMD is used to decompose the time series into a series of IMFs (intrinsic mode functions), each representing a different scale or frequency component of the data. The NARDL model is then applied to each IMF to capture the nonlinear relationships between the variables. The MNARDL (multiscaled NARDL) model is shown to outperform other existing models in terms of accuracy, particularly in cases where the data exhibits nonlinearities and non-stationarities. The model is demonstrated through simulation studies and real-world applications in the areas of economics and finance. The article presents a novel approach to time series forecasting that integrates two existing techniques and provides improved accuracy for modeling nonlinear data.

This paper [19] explores the challenges of predicting stock market movements, a key area of interest across var-

ious fields such as statistics, AI, and finance. It highlights the importance of accurate predictions in reducing investment risks and emphasizes that machine learning models often outperform traditional statistical approaches. Specifically, the study investigates the Gaussian Naïve Bayes (GNB) algorithm, which has not been extensively studied in this context. The researchers evaluate GNB's performance when integrated with different feature scaling and extraction techniques, using Kendall's test of concordance for ranking. Results indicate that the GNB model combined with Linear Discriminant Analysis (GNB_LDA) outperformed other models in three of four evaluation metrics (accuracy, F1-score, and AUC). Additionally, the GNB model using Min-Max scaling and PCA achieved the highest specificity rank, demonstrating that GNB performs better with Min-Max scaling than with standardization techniques.

Authors of the article [8] discuss the challenges of predicting foreign exchange rates due to their complex and volatile nature. The paper proposes a new model that combines two powerful neural networks, the GRU (Gated Recurrent Unit) and LSTM, to predict the future closing prices of four major currency pairs. The proposed hybrid model outperforms standalone LSTM and GRU models, as well as a simple moving average-based statistical model, for a 10-minute timeframe and provides the best result for two currency pairs in terms of MSE (Mean Square Error), RMSE (Root Mean Square Error), and MAE (Mean Absolute Error) performance metrics for a 30-minute timeframe. The model's performance is validated using MSE, RMSE, MAE, and R score, with the proposed hybrid GRU-LSTM model proving to be the least risky among all compared models in terms of R2 score.

Forecasting fast and high-frequency financial data is challenging due to the dynamic and chaotic nature of stock markets. This study [4] presents a novel hybrid model combining fractional order derivatives and deep learning, specifically LSTM networks, to predict sudden market fluctuations. Traditional methods like data mining and statistical approaches struggle with stock price variability, but the proposed ARFIMA-LSTM (Autoregressive fractionally integrated moving average LSTM) hybrid model effectively extracts features and models non-linear functions. It overcomes volatility and overfitting issues, outperforming traditional methods by achieving approximately 80% accuracy improvement in RMSE when evaluated with real stock market data from the PSX company.

The article [1] investigates the relationship between public debt and economic growth in Morocco. The authors use an ARDL model to analyze the impact of public debt on economic growth, taking into account other factors such as investment, exports, and inflation. They use data from 1980 to 2019 to estimate the model. The results show that there is a negative relationship between public debt and economic growth in Morocco, indicating that an increase in public debt can lead to a decrease in economic growth. The authors also find that investment and exports have a positive

impact on economic growth, while inflation has a negative impact. The article highlights the importance of managing public debt in order to promote economic growth in Morocco. The findings may be useful for policymakers in the country as they make decisions about fiscal policy and debt management.

Behavioral finance studies suggest that emotions impact stock markets. This paper [3] discusses a method to collect and analyze sentiment from various sources about Casablanca Stock Exchange. Using sentiment analysis and machine learning, it aims to link public sentiment to stock market performance.

The paper [18] presents a financial risk prediction model utilizing graph networks to address inaccuracies in enterprise financial risk and profit predictions. It integrates multi-scale feature extraction and sequence analysis, employing a bidirectional gated recurrent unit to effectively capture temporal relationships in time series data. The profit prediction model combines multi-scale advantages with attention mechanisms to enhance the identification of influential features, significantly improving predictive accuracy. After iterative training, the model achieved an accuracy of 98.03% and an F1 score of 0.98 for financial risk predictions. The profit prediction model outperformed others in regression and classification metrics, with a mean square error of just 0.0232. Overall, both models demonstrate strong predictive capabilities and practical significance.

This article [16] discusses the increasing importance of electric vehicle load scheduling in grid power scheduling due to the rising use of electric vehicles. The effective electric vehicle power dispatching helps balance the peak-valley difference of power dispatching, increase the power supply utilization rate, and reduce the power supply pressure of line transformer. The article summarizes the research status of electric vehicle charging load, analyzes traditional charging load research methods and proposes a charging load forecasting method combining XGBoost and LSTM. The proposed method is based on the prediction results of the XGBoost model for feature engineering and statistical methods, and training the LSTM model for load prediction. The charging station load forecasting method studied in this paper can support the regional load forecasting research of electric vehicles with high permeability and further optimize power dispatching. The proposed method is verified using the data of a charging station in Jiangsu.

This paper [7] introduces a new approach to predict future returns in volatile and nonlinear financial markets. It consists of three stages: fractal modeling and recurrence analysis, Granger causality tests, and wavelet transformation. Machine learning algorithms are applied to learn patterns and make predictions. Testing with Asian emerging stock indexes from 2012 to 2017 shows that this framework is effective for forecasting.

The authors of [9] discuss the challenge of forecasting stock prices due to the volatile nature of the stock market, which makes it difficult to apply linear models or simple

time-series or regression techniques. The author suggests that SVM (support vector machine) is a good alternative tool for stock forecasting, as it is a popular machine learning technique for the capital investment industry that can forecast financial data more accurately. The article presents an experiment that examines the stock prices of 5 Moroccan banks and shows that SVM can perform better when the global evolution of the market is added to the independent variables. To express the global evolution of the market, the author uses three indices of the Casablanca Stock Exchange: MASI, MADEX (Moroccan Most Active Shares Index), and Banks Sector Index. Also, this article highlights the potential of SVM for stock price prediction, and emphasizes the importance of considering the global market conditions as a variable to improve forecasting accuracy. The findings may be useful for investors and financial analysts looking for new methods to improve their stock trading and investment decisions. Below is a comparative table (Table 1) of the results of the related work, with their results.

In conclusion, these studies provide evidence that LSTM, XGBoost, and ARDL models can all be effectively used to forecast exchange rates. The results suggest that LSTM is generally better in terms of accuracy and stability, while XGBoost is faster in training and prediction time. In this current work, we aim to demonstrate that ARDL models have distinct advantages, particularly in the context of the Moroccan market, as they take into account exogenous variables during events and can be explained with a formula, contrasting with the black-box nature of machine learning algorithms. The unique behavior of the Moroccan market necessitates a deeper exploration of these tailored approaches.

3 Methodology

The article aims to explore the possible short-term and long-term relationships between the COVID-19 pandemic and the Moroccan financial market. The study also investigates the potential of the ARDL model in improving the accuracy of future value predictions for the Moroccan financial market, compared to LSTM and XGBOOST.

3.1 COVID19 impact on Moroccan financial market based on ARDL

Most financial market studies use VAR (Vector autoregression) modeling to analyze the relationship between explanatory and explained variables. However, this method requires that the series be integrated of the same order, which is not always the case for macroeconomic series. To address this issue, Pesaran, Shin, and Smith (2001) proposed the ARDL method, which considers the limitations of the VAR model. This approach, which tests the long-

run relationship based on variables with different integration orders, is an alternative to cointegration tests. The ARDL method provides unbiased estimates of the long-run relationship and is more suitable for small samples. In this study, we will use the ARDL model to investigate the short- and long-run relationships between COVID-19 pandemic and Moroccan financial market. We will also determine the optimal number of lags using AIC (Akaike Information Criterion) and test for the presence of causal relationships between the variables.

3.1.1 Research model

The application of the ARDL model, represented by the subsequent equation, will enable the estimation of the responses to the hypotheses stated underneath :

$$\Delta \text{LogMASI}_{(t)} = C + \sum_{i=1}^p \alpha_{1i} \Delta \text{LogMASI}_{(t-i)} + \sum_{i=0}^q \alpha_{2i} \Delta \text{LogCOVID}_{(t-i)} + \beta_1 \text{LogMASI}_{(t-1)} + \beta_2 \text{LogCOVID}_{(t-1)} + \varepsilon(t) \quad (1)$$

(Table 2) provides a description of the various variables included in the equation.

Table 2: Description of equation variables

Variable	Description
MASI	Price of the Moroccan All Shares Index
COVID	Daily new cases of COVID-19
C	Constant
Log()	Natural logarithm operator
Δ	First difference operator
α ₁ ; α ₂	Short-run coefficients
β ₁ ; β ₂	Long-run dynamics
ε(t)	Error term

It is important to note that before estimating the ARDL model, a cointegration test must be performed. This is because it is not possible to estimate an error correction model or determine the short and long-term effects for variables that are not cointegrated. In the case of long-term effects, we conduct a cointegration limit test based on the Fisher statistic, with the hypothesis being that the variables are cointegrated.

$$H_0 : \beta_1 = \beta_2 = 0$$

If H_0 is rejected, it indicates the existence of cointegration. The F-statistic should surpass the upper bounds of $I(1)$ to reject H_0 , but if it is less than the lower critical bounds of $I(0)$, H_0 is accepted. Otherwise, no conclusion can be made. Once cointegration is confirmed, the long-term relationship is determined by eliminating the variables in first difference (Morley, 2006 and Antoniou et al., 2013). Utilizing equation (1), we can deduce that the relationship is illustrated by the subsequent equation :

Table 1: Summary of studies and their findings

Work	Models	Dataset	Results
[6]	ARDL	Saudi Arabia (TASI)	The study finds a long-term negative relationship between LOG_TASI and LOGCOVID_19, with unidirectional causality from COVID-19 to TASI, highlighting the need for strong national measures to prevent a significant stock market crash in KSA.
[14]	panel-EGLS and panel quantile regression approaches	Different emerging markets	Finding that new cases and deaths negatively influence returns, while increased testing positively affects them, with some variations between developed and emerging markets.
[10]	ARDL	Bangladesh market	The study revealed that financial innovation positively influences economic growth both in the short and long run, which is statistically significant as well.
[13]	Nonlinear Panel ARDL	Oil-stock nexus	The study finds that incorporating positive and negative oil price changes improves stock price forecasts only for oil-importing countries, highlighting a key difference in the oil price-stock relationship between importing and exporting nations.
[12]	Multiscaled Neural ARDL	Oil & Bitcoin	The empirical experiments conducted on real-world economic data prove that the decomposing framework significantly improves the forecasting accuracy.
[8]	GRU-LSTM hybrid network	EUR/USD, GBP/USD, USD/CAD	The hybrid GRU-LSTM model outperforms standalone LSTM, GRU, and SMA models in predicting GBP/USD and USD/CAD currency pairs, demonstrating the best performance metrics and lowest risk overall.
[1]	ARDL	GDP	The study using an Auto Regressive Distributed Lag model reveals that total government debt significantly negatively impacts economic growth in Morocco both in the short term (0.62% decrease in growth for a 1% increase in debt) and long term (0.4% decrease), while the investment rate positively influences growth; however, bank credit to the private sector remains statistically insignificant, highlighting challenges in the trade balance and the need for broader economic reforms.
[9]	SVM	MASI & MADEX	SVM significantly improves stock forecasting accuracy for five Moroccan banks, especially when incorporating global market trends like the MASI, MADEX, and Banks Sector Index.

$$\text{LogMASI}_{(t)} = - \left(\frac{C}{\beta_1} \right) - \left(\frac{\beta_2}{\beta_1} \right) \text{LogCOVID}_{(t)} \quad (2)$$

The confirmation of the presence or absence of cointegration between the variables can be done by using ECM (error correction model) for eq.2 as demonstrated below:

$$\begin{aligned} \Delta \text{LogMASI}_{(t)} = & \sum_{i=1}^p \alpha_{1i} \Delta \text{LogMASI}_{(t-i)} \\ & + \sum_{i=0}^q \alpha_{2i} \Delta \text{LogCOVID}_{(t-i)} \\ & + \beta_1 \text{ECM}_{(t-1)} + \varepsilon_t \end{aligned} \quad (3)$$

$$\begin{aligned} \text{ECM}_{(t)} = & \text{LogMASI}_{(t)} - \\ & \left[- \left(\frac{C}{\beta_1} \right) - \left(\frac{\beta_2}{\beta_1} \right) \text{LogCOVID}_{(t)} \right] \end{aligned} \quad (4)$$

In the ARDL model, trend and seasonality are incorporated through lagged values of MASI index prices and COVID-19 cases, selected to capture short-term fluctuations and long-term equilibrium adjustments. The trend component is addressed by using previous levels of the MASI, reflecting persistent market shifts, while seasonality is captured by lags that align with observed recurring patterns in COVID-19 data. This structure ensures the model can respond to both immediate impacts and ongoing trends in the market, particularly during volatile periods like the COVID-19 pandemic.

3.2 Forecasting MASI index with ARDL, LSTM and XGBoost

We hypothesize that by using ARDL, LSTM and XGBoost models, we can accurately forecast the future values of the MASI. Specifically, we expect that:

- The ARDL model will perform well in predicting MASI since it can capture both short and long-run dynamics of the data. We expect that the model will be able to identify significant relationships between MASI and other relevant economic indicators, such as COVID19 cases, inflation rate, and interest rate.
- The LSTM model will also perform well in predicting MASI since it can capture complex temporal dependencies in time series data. We expect that the model will be able to learn and identify the patterns in MASI and its determinants over time, thereby improving its forecasting accuracy.
- The XGBoost model will perform well in predicting MASI since it is an ensemble tree-based method that can capture non-linear relationships and interactions between variables. We expect that the model will be able to identify the most important features that influence MASI, and thereby provide more accurate forecasts than traditional regression-based models.

4 Results and discussion

4.1 Data and description

Our study focuses on analyzing the impact of the COVID19 pandemic on the MASI, using daily closing prices that were obtained from www.investing.com. We also collected data on the daily number of confirmed COVID19 cases from the official website of the Moroccan Ministry of Health. The data covers the period from March 03, 2020 (the day when the first COVID19 case was reported in Morocco) to February 11, 2022. Prior to analysis, both variables were subjected to a log transformation. (the data is accessible via this link <https://osf.io/umjgb/files>).

4.2 Descriptive statistics

The following table (Table 3) displays the descriptive statistics for two variables analyzed in our study. The first variable, MASI, represents the closing price of MASI, while the second variable, NEW_CASES, represents the number of new daily confirmed cases of COVID-19 in Morocco.

Table 3: Descriptive statistics of MASI and COVID-19 new cases (2 March 2020 to 11 February 2022)

	MASI	NEW_CASES
Mean	11524.53	1576.031
Median	11517.92	579.0000
Maximum	13991.47	12039.00
Minimum	8987.890	0.000000
Std. Dev.	1347.250	2219.878
Skewness	0.018541	2.233916
Kurtosis	1.856586	8.076603
Jarque-Bera	26.44808	924.1962
Observations	485	485

The two figures (1,2) show the trends of MASI price and COVID19 new cases over the same period.

4.3 ARDL quantitative results and discussions

4.3.1 Stationarity (Unit root tests)

This part discusses the concept of non-stationarity in time series data, where a series with a moving average and/or variance that varies over time is considered non-stationary. If not addressed through stationarization, this non-stationarity can lead to "spurious" regressions. To determine if a series is stationary or not (i.e., if a unit root exists), several tests can be used, such as ADF (augmented Dickey-Fuller) test, PP (Phillippe-Perron) test, AZ (Andrews and Zivot) test, the Ng test-Perron, and KPSS (Kwiatkowski-Phillips-Schmidt-Shin) test. Among these tests, the ADF and PP tests are the most commonly used and easy to apply. The ADF test is effective in cases of autocorrelation of errors, while the PP test is suitable in the

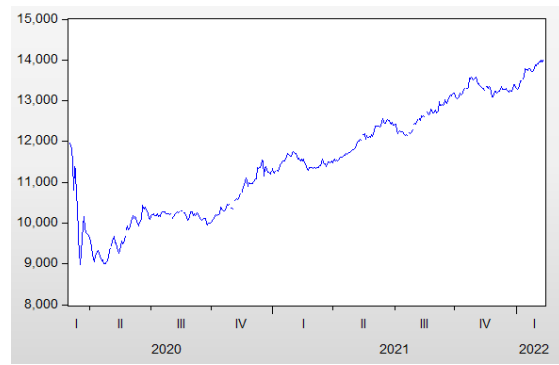


Figure 1: Movement of MASI price in the concerned period

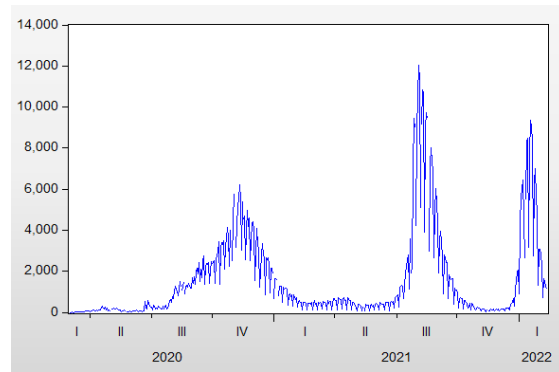


Figure 2: COVID19 new cases in the concerned period

presence of heteroscedasticity. In this study, the ADF and PP tests were used with the test critical values of Mackinnon (1996), and the results are presented in (Table 4).

The results of the study show that the Log MASI Level data is non-stationary, while the Log MASI 1st difference and Log COVID19 data are stationary. The optimal number of lags is 17 based on AIC, and all variables have a significance at the 1% level. The ADF test statistic values are also reported for each variable, with the Log MASI Level having a test statistic value greater than the critical value and a p-value greater than 0.05, indicating non-stationarity. Meanwhile, the Log MASI 1st difference and Log COVID19 have test statistic values lower than their respective critical values and p-values lower than 0.05, indicating stationarity (Table 5).

4.3.2 ARDL estimation

To select the most optimal ARDL model with statistically significant results and fewer parameters, we utilize AIC. The model is estimated with the "constant & trend" option because of its high significance (Probability < 1%). The following are the estimation results of the selected optimal ARDL model.

The optimal model is selected based on the AIC value, where the model with the smallest AIC value (Figure 3) is considered the best. In this study, the optimal model is the ARDL(2,6), which is statistically significant with a global

Table 4: ADF Unit Root test on the log level of variables

Variables	Level					Integration order
	T-statistic	1%	5%	10%	P-value	
Log MASI	-0.8291	-3.4441	-2.8675	-2.5700	0.8095	I(1)
Log MASI 1st Difference	-5.3066	-3.4441	-2.8675	-2.5700	0.0000	
Log COVID	-3.0442	-3.4442	-2.8675	-2.5700	0.0317	I(0)

Table 5: PP Unit Root test on the log level of variables

Variables	Level					Integration order
	T-statistic	1%	5%	10%	P-value	
Log MASI	-0.3695	-3.4436	-2.8672	-2.5698	0.9114	I(1)
Log MASI 1st Difference	-18.1950	-3.4436	-2.8673	-2.5699	0.0000	
Log COVID	-5.6379	-3.4437	-2.8673	-2.5699	0.0000	I(0)

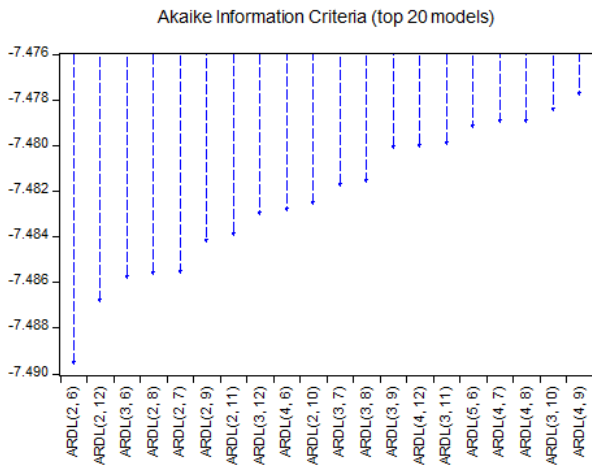


Figure 3: Akaike information criteria

Prob (F-statistic) value of 0.0000 (Table 6).

While most of the coefficients in the model are significant, it is necessary to conduct validity tests such as autocorrelation tests to ensure its validity (Table 7). Additionally, the model is globally significant.

Based on the Ljung-Box test results, the Q-statistic probability is above the 5% and 10% thresholds for all results, indicating the absence of autocorrelation in the model errors. This is important because the presence of autocorrelation in residuals can lead to inconsistent parameter estimates due to the inclusion of a lagged dependent variable as an exogenous variable in the model.

The F test value of 9.43 (Table 8) exceeds the majority of the I(1) bounds, indicating significant **cointegration** between the variables at a 2.5% significance level. This suggests that it is possible to estimate the long-term effects of Log_COVID19 on Log_MASI.

The long term relationship is described as follows : $Log_MASI = -0.0020 * Log_COVID$. The results (Table 9) show that there is a significant negative long-term relationship between COVID 19 and the stock market in Morocco: a 100% increase in the daily number of confirmed cases of

COVID-19 resulted in a 0.2% decrease in the MASI price. In the short-term relationship, it appears that there is no significance between all variables, but Log_COVID delayed by 5 days (t-5) has a positive impact on Log_MASI in day (t) at the 10% level.

Since correlation does not necessarily imply causality, we must test the causality that may exist between the variables, we use the Toda Yamamoto causality test.

A causal relationship from Log_COVID to Log_MASI is confirmed by the Toda-Yamamoto causality test (Prob = 0.0005, the null hypothesis is rejected) (Table 10). However, there is no causality between Log_MASI and Log_COVID (Pob=0.0857).

4.4 Forecasting results

4.4.1 Performance measure

The MAPE (Mean Absolute Percentage Error) is a practical metric used to evaluate the accuracy of forecasting models. It calculates the average percentage deviation of forecasted values from the observed values. By expressing the error as a percentage, the MAPE enables easy comparison between different models. The formula for MAPE is as follows:

$$MAPE = 100\% - \frac{100\%}{n} \sum_{t=1}^n \left| \frac{Forecasted_value_t}{Real_value_t} \right| \tag{5}$$

With **n** the number of forecasted values.

4.4.2 Results

(Table 11) outlines the various approaches utilized to forecast the MASI index (ARDL, LSTM and XGBOOST), along with the corresponding inputs for each method.

A greedy algorithm was applied to determine the optimal lag values for the ARDL model, focusing on selecting the lagged variables that best capture the short-term dynamics of the MASI index. Specifically, the model uses a unique set of inputs based on two-day lags for MASI prices and six-day lags for new confirmed COVID-19 cases. This selection reflects an iterative process in which different lag

Table 6: ARDL estimation

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
MASI_LOG(-1)	1.196800	0.043909	27.25662	0.0000
MASI_LOG(-2)	-0.257500	0.043916	-5.863430	0.0000
COVID19_LOG	0.001027	0.000702	1.461975	0.1444
COVID19_LOG(-1)	-0.000982	0.000711	-1.382608	0.1675
COVID19_LOG(-2)	0.000570	0.000559	1.020708	0.3079
COVID19_LOG(-3)	-0.001067	0.000566	-1.885020	0.0601
COVID19_LOG(-4)	-0.000443	0.000557	-0.795811	0.4265
COVID19_LOG(-5)	-0.000417	0.000710	-0.586752	0.5577
COVID19_LOG(-6)	0.001191	0.000681	1.749257	0.0809
C	0.555984	0.128045	4.342094	0.0000
@TREND	5.14E - 05	1.22E - 05	4.212603	0.0000
R-squared	0.997024	Mean dependent var		9.346391
Adjusted R-squared	0.996960	S.D. dependent var		0.118471
S.E. of regression	0.006532	Akaike info criterion		-7.201220
Sum squared resid	0.019756	Schwarz criterion		-7.104652
Log likelihood	1717.689	Hannan-Quinn criter.		-7.163241
F-statistic	15512.26	Durbin-Watson stat		2.022289
Prob(F-statistic)	0.000000			

*Note: p-values and any subsequent tests do not account for model selection.

Table 7: Autocorrelation of residuals.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*
. .	. .	1 -0.031	-0.031	0.4672	0.494
. .	. .	2 0.030	0.029	0.8915	0.640
. .	. .	3 -0.059	-0.057	2.5576	0.465
. .	. .	4 -0.055	-0.060	4.0298	0.402
. *	. *	5 0.082	0.082	7.2305	0.204
. .	. .	6 0.020	0.026	7.4317	0.283
. *	. *	7 0.099	0.090	12.155	0.096
. .	. .	8 -0.029	-0.019	12.565	0.128
. .	. .	9 -0.019	-0.015	12.741	0.175
. .	. .	10 -0.038	-0.032	13.453	0.199

*Probabilities may not be valid for this equation specification.

Table 8: Bound test to cointegration results

F-Bounds Test			Null Hypothesis: No levels relationship	
Test Statistic	Value	Signif.	Lower bound I(0)	Upper bound I(1)
Asymptotic: n=1000				
F-statistic	9.433081	10%	5.59	6.26
k	1	5%	6.56	7.3
		2.5%	7.46	8.27
		1%	8.74	9.63

Table 9: Dynamics of the short-run and long-run

Conditional Error Correction Regression				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
MASI_LOG(-1)*	-0.060700	0.013991	-4.338372	0.0000
COVID19_LOG(-1)	-0.000121	0.000223	-0.545517	0.5857
D(MASI_LOG(-1))	0.257500	0.043916	5.863430	0.0000
D(COVID19_LOG)	0.001027	0.000702	1.461975	0.1444
D(COVID19_LOG(-1))	0.000166	0.000680	0.243598	0.8077
D(COVID19_LOG(-2))	0.000736	0.000706	1.042008	0.2980
D(COVID19_LOG(-3))	-0.000331	0.000706	-0.469430	0.6390
D(COVID19_LOG(-4))	-0.000774	0.000677	-1.144776	0.2529
D(COVID19_LOG(-5))	-0.001191	0.000681	-1.749257	0.0809
EC = Log_MASI - (-0.0020*COVID19_LOG)				
Long-Run Coefficients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
COVID19_LOG	-0.002001	0.003616	-0.553284	0.5803
C	0.555984	0.128045	4.342094	0.0000
@TREND	5.14E - 05	1.22E - 05	4.212603	0.0000

Table 10: Toda-Yamamoto causality test.

Dependent variable: MASI_LOG				
Excluded	Chi-sq	df	Prob.	
COVID19_LOG	15.32437	2	0.0005	
All	15.32437	2	0.0005	
Dependent variable: COVID19_LOG				
Excluded	Chi-sq	df	Prob.	
MASI_LOG	4.914155	2	0.0857	
All	4.914155	2	0.0857	

Table 11: Forecasting results

Method	Inputs	MAPE	Processing time	
			Training	Forecasting
ARDL	Lags only	34.4%	1s	
ARDL	Lags + trend	32.3%	1s	
ARDL	Lags + seasonality	32.1%	1s	
ARDL	Lags + trend + seasonality	26.7%	1s	
LSTM	Lags	30.5%	2min 56s	21s
XGBOOST	Lags	32%	6min	1s

combinations were tested, with the greedy algorithm identifying the combination that minimized the forecasting error. By emphasizing shorter lags, this approach aims to model the immediate impacts of recent MASI price fluctuations and COVID-19 case data, as these factors are expected to influence market movements in the short term. This method not only simplifies the model but also enhances its responsiveness to recent changes, thereby improving predictive

accuracy.

In contrast, LSTM and XGBOOST models rely on a grid search of hyperparameters to select the best combination of variables for the forecast. This method involves testing different combinations of input variables to determine the most optimal set for predicting the MASI index accurately.

After evaluating the performance of all three methods, the ARDL model with lags, trend, and seasonality variables

outperformed both LSTM and XGBOOST models in terms of accuracy and processing time. This finding indicates that the inclusion of trend and seasonality variables in addition to the lags of MASI prices and new confirmed cases data significantly improves the model's accuracy.

The ARDL model's execution time of only 1 second is also impressive and demonstrates its efficiency, particularly when compared to the relatively more computationally expensive LSTM and XGBOOST models. Overall, the ARDL model with its unique set of inputs, outperforms the LSTM and XGBOOST models in terms of accuracy, efficiency, and computational resources.

4.4.3 Forecasting discussion

In this section, we compare the forecasting results of the ARDL, LSTM, and XGBoost models with findings from existing literature, shedding light on why the ARDL model outperformed the others and analyzing the conditions that contributed to its superior performance.

The ARDL model demonstrated a remarkable accuracy, achieving a MAPE of 26.7% when incorporating lags, trend, and seasonality variables. This performance is particularly noteworthy when juxtaposed with the LSTM and XGBoost models, which recorded MAPEs of 30.5% and 32%, respectively. A primary reason for the ARDL model's success lies in its ability to effectively handle short-term dynamics using a specific set of inputs—namely, the lags of MASI prices and new confirmed COVID-19 cases. These variables are likely to exert an immediate impact on the market, allowing ARDL to capture the temporal relationships more adeptly.

The findings align with existing literature that emphasizes the efficacy of ARDL in time series forecasting, particularly in scenarios characterized by smaller sample sizes. The simplicity of the ARDL model, combined with its explicit incorporation of exogenous variables, allows for easier interpretation and better forecasting accuracy under conditions where non-linear relationships might not be as pronounced. In contrast, both LSTM and XGBoost, while powerful in handling complex, non-linear patterns, require larger datasets to truly exploit their capabilities. In our case, the relatively limited data available during the COVID-19 period may have hindered these models from achieving optimal performance.

Additionally, the computational efficiency of the ARDL model—requiring only 1 second for execution—highlights its practicality for real-time forecasting, particularly in fast-moving markets like the Moroccan stock market. In contrast, LSTM's training time of 2 minutes and 56 seconds, along with a forecasting time of 21 seconds, and XGBoost's 6 minutes for training, demonstrate the trade-off between model complexity and computational demand.

This study's results suggest that while machine learning models like LSTM and XGBoost offer sophisticated techniques for capturing non-linear patterns, the unique conditions of the Moroccan market, combined with the character-

istics of the dataset, made ARDL the more suitable choice for this specific forecasting task. Future work could explore hybrid models that combine the strengths of ARDL with machine learning techniques, potentially leading to enhanced accuracy. Additionally, further research should investigate the scalability of these findings to other emerging markets with similar characteristics, as well as the implications of larger datasets that might better inform machine learning approaches.

4.4.4 Computational efficiency discussion

The ARDL model is the most computationally efficient in both memory usage and scalability due to its simple linear regression structure. It requires minimal memory since it only stores a limited number of lagged values and coefficients, making it ideal for applications needing quick and efficient forecasts on moderate datasets. However, its simplicity may limit its performance with larger or more complex datasets.

In contrast, the LSTM model, while powerful in capturing complex, long-term dependencies in sequential data, is much more memory-intensive due to its multi-layered architecture and need to store information across each time step. This memory requirement restricts its scalability unless specialized hardware like GPUs is used, making LSTM better suited for smaller datasets or scenarios where memory resources are abundant.

XGBoost balances efficiency and scalability well. It requires more memory than ARDL due to its ensemble of decision trees, but it is significantly more scalable because of its parallel processing capabilities. Optimized for handling large datasets and sparse data, XGBoost is ideal for applications prioritizing accuracy on large datasets, although it requires moderate memory availability for efficient processing. Overall, ARDL offers the most efficient option for smaller datasets, while XGBoost and LSTM trade off memory and computational resources for accuracy in large and complex data scenarios.

4.5 Summary and conclusions

This study focuses on modeling the impact of the Covid-19 pandemic on the Moroccan stock market using the ARDL estimation approach. The study analyzes both short-term and long-term relationships between the MASI index and the number of daily confirmed Covid-19 cases, indicating a negative long-term relationship and unidirectional causality from Covid-19 to the MASI index.

To capture the short-term dynamics of the MASI index, the ARDL model uses a unique set of inputs based on lags of MASI prices for two previous days and new confirmed cases data from the previous six days. In contrast, LSTM and XGBOOST models use a grid search of hyperparameters to select the optimal set of input variables for accurate forecasting.

After evaluating the performance of all three models,

the ARDL model with lags, trend, and seasonality variables outperforms both LSTM and XGBOOST models in terms of accuracy and processing time. The inclusion of trend and seasonality variables significantly improves the model's accuracy, and the ARDL model's execution time of only 1 second demonstrates its efficiency compared to the relatively more computationally expensive LSTM and XGBOOST models. In summary, the ARDL model with its unique set of inputs proves to be the best option for accurately forecasting the MASI index during the Covid-19 pandemic.

The ARDL model assumes a stable, long-term relationship between variables, requiring that the data series be either stationary or integrated of the same order (usually $I(0)$ or $I(1)$). To meet this, unit root tests are performed before model estimation. Additionally, ARDL presumes no perfect multicollinearity among explanatory variables, ensuring independent impacts, and it assumes that residuals are normally distributed and homoscedastic—meaning constant variance of error terms over time. If these assumptions are not met, the coefficient estimates may be biased or inefficient, impacting inference validity. Thus, diagnostic tests for heteroscedasticity, normality, and autocorrelation are conducted post-estimation to confirm the model's reliability.

The results offer actionable insights that could guide trading and investment strategies by providing a reliable approach to forecasting market trends in the short term, especially under fluctuating conditions such as those influenced by COVID-19 cases. For investors, the ARDL model's accuracy and efficiency in processing time may enhance their ability to make timely, data-driven decisions, thereby improving portfolio performance. Policymakers could also leverage these insights to better understand market responses to economic shocks or health crises, enabling more informed policy adjustments that help stabilize or stimulate the financial sector. Overall, the study contributes valuable tools that can support more informed decision-making across different roles in the Moroccan financial market.

5 List of abbreviations

ARDL	AutoRegressive Distributed Lag
LSTM	Long Short-Term Memory
XGBOOST	Extreme gradient boosting
MASI	Moroccan All Shares Index
EMD	Empirical mode decomposition
NARDL	Neural autoregressive distributed lag
IMFs	Intrinsic mode functions
MARDL	Multiscaled NARDL
GRU	Gated Recurrent Unit

MSE	Mean Square Error
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
ARFIMA	Autoregressive fractionally integrated moving average
SVM	Support Vector Machine
MADEX	Moroccan Most Active Shares Index
VAR	Vector autoregression
AIC	Akaike Information Criterion
ECM	Error correction model
ADF	Augmented Dickey-Fuller
PP	Phillippe-Perron
AZ	Andrews and Zivot
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
MAPE	Mean Absolute Percentage Error

References

- [1] Achibane, Khalid. *Study of the impact of Public Debt on Moroccan Economic Growth: ARDL model*. International Journal of Accounting, Finance, Auditing, Management and Economics, pages 460–472, 2021.
- [2] Ashraf, Badar Nadeem. *Economic impact of government interventions during the COVID-19 pandemic: International evidence from financial markets*. Journal of Behavioral and Experimental Finance, 100371, 2020.
- [3] Bourezk, Hind; Raji, Amine; Acha, Nawfal; Barka, Hafid. *Analyzing Moroccan Stock Market using Machine Learning and Sentiment Analysis*. 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)T, pages 1–5, 2020.
- [4] Bukhari, Ayaz Hussain; Raja, Muhammad Asif Zahoor; Sulaiman, Muhammad; Islam, Saeed; Shoaib, Muhammad; Kumam, Poom. *Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting*. Journal of Physics: Conference Series, pages 71326-71338, 2020.
- [5] Cerniglia, Joseph; Fabozzi, Frank. *Selecting Computational Models for Asset Management: Financial Econometrics versus Machine Learning—Is There a Conflict?*. The Journal of Portfolio Management, 47, 2020.

- [6] Chaouachi, Maroua; Chaouachi, Slim. *Current covid-19 impact on Saudi stock market: Evidence from an ARDL model*. Available at SSRN 3636333, 2020.
- [7] Ghosh, Indranil; Jana, Rabin K.; Sanyal, Manas K. *Analysis of temporal pattern, causal interaction and predictive modeling of financial markets using non-linear dynamics, econometric models and machine learning algorithms*. Applied Soft Computing, pages 105553, 2019.
- [8] Islam, Md Saiful; Hossain, Emam. *Foreign exchange currency rate prediction using a GRU-LSTM hybrid network*. Soft Computing Letters, pages 100009, 2021.
- [9] NAHIL, Anass; Lyhyaoui, Abdelouahid. *Stock price prediction based on SVM: The impact of the stock market indices on the model performance*. Proceedings of the Engineering and Technology–PET, pages 91–95, 2017.
- [10] Qamruzzaman, Md; Wei, Jianguo. *Financial innovation, stock market development, and economic growth: An application of ARDL model*. International Journal of Financial Studies, 69, 2018.
- [11] Rhif, Manel; Ben Abbes, Ali; Farah, Imed Riadh; Martínez, Beatriz; Sang, Yanfang. *Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review*. Applied Sciences, 7, 2019.
- [12] Saâdaoui, Foued; Ben Messaoud, Othman. *Multi-scaled neural autoregressive distributed lag: A new empirical mode decomposition model for nonlinear time series forecasting*. International Journal of Neural Systems, pages 2050039, 2020.
- [13] Salisu, Afees A; Isah, Kazeem O. *Revisiting the oil price and stock market nexus: A nonlinear Panel ARDL approach*. Economic Modelling, 258–271, 2017.
- [14] Ullah, Sabeeh. *Impact of COVID-19 pandemic on financial markets: A global perspective*. Journal of the Knowledge Economy volume, 982–1003, 2023.
- [15] Weng, Bin; Ahmed, Mohamed A.; Megahed, Fadel M. *Stock market one-day ahead movement prediction using disparate data sources*. Expert Systems with Applications, 153-163, 2017.
- [16] Xue, Mingfeng; Wu, Lin; Zhang, Qi Pei; Lu, Ji Xi-ang; Mao, Xiaobo; Pan, Yongtao. *Research on load forecasting of charging station based on XGBoost and LSTM model*. Journal of Physics: Conference Series, pages 012145, 2021.
- [17] Zaghum Umar; Onur Polat; Sun-Yong Choi; Tamara Teplova. *The impact of the Russia-Ukraine conflict on the connectedness of financial markets*. Finance Research Letters, 102976, 2022.
- [18] Liu, Yangyan and Pan, Bolin *Profit Estimation Model and Financial Risk Prediction Combining Multi-scale Convolutional Feature Extractor and BGRU Model*. Informatica, 48, 2024.
- [19] Ampomah, Ernest Kwame and Nyame, Gabriel and Qin, Zhiguang and Addo, Prince Clement and Gyamfi, Enoch Opanin and Gyan, Micheal *Stock market prediction with gaussian naïve bayes machine learning algorithm*. Informatica, 45, 2021.