

ARIMA-Based Forecasting of S&P BSE SENSEX Returns

Deep Dutta

Research Scholar, University of Calcutta, West Bengal, India

Corresponding Author: deepduttacoc@gmail.com

Received: 03-11-2023

Revised: 19-11-2023

Accepted: 02-12-2023

ABSTRACT

Investment in the stock market requires a delicate balance between profitability and risk management, with risk aversion playing a vital role. This study explores the ARIMA forecasting method to predict S&P BSE SENSEX returns, providing valuable insights for investors and financial experts. Using a 3-year dataset, the ARIMA (3,1,1) model was identified as the optimal choice. Diagnostic checks confirmed its reliability, ensuring unbiased and accurate forecasts. In static forecasting, the model exhibited high-quality performance with low error rates. Dynamic forecasting further revealed precision in predicting future values. While the ARIMA model aids in making informed financial decisions, it's crucial to acknowledge its limitations. This research contributes to the understanding of stock market forecasting methodologies, benefiting investors and analysts in navigating this dynamic landscape.

Keywords: arima, forecasting, s&p bse sensex

I. INTRODUCTION

Investment in the stock market necessitates the achievement of an optimal equilibrium between profitability and risk. To accomplish this objective, a comprehensive understanding of market dynamics, particularly in the realm of risk prediction and management, is indispensable. The notion of risk aversion is of great significance to various stakeholders, including investors, policymakers, researchers, and financial experts, due to its impact on portfolio diversification and market stability.

The interplay between investment returns and risks exerts a significant influence on decision-making processes. Successful investors strive to transform each action into substantial returns, relying on effective and rational strategies, as underscored by Kaufman (1995). Empirical research offers evidence supporting a positive correlation between stock markets and economic growth (Guptha & Rao, 2018; Kim et al., 2011; Mallikarjuna & Rao, 2019), which underscores the crucial role of investment decisions in attaining desired financial outcomes.

Nevertheless, stock markets are inherently characterized by their dynamic, intricate, and volatile nature. Predicting stock prices and returns in such an environment poses a formidable challenge. In this context, a delve into the realm of ARIMA (Auto-Regressive Integrated Moving Average) forecasting, a robust analytical tool that offers valuable insights into the future movements of the S&P BSE SENSEX, thereby providing assistance to investors as they navigate the ever-changing landscape of the stock market.

II. LITERATURE REVIEW

In a recent study conducted by Neely et al. (2014), the authors emphasized the significance of employing technical indicators to forecast stock returns, illustrating their relevance from both an economic and statistical perspective. These findings align with a wider body of research that has investigated the predictability of stock returns, as demonstrated by studies conducted Zhu & Zhu (2013) by and Jiahan & Ilias, (2017).

Chari & Henry (2004) provided valuable insights into the reduction of systematic risk in stock market liberalizations. They argued that as the global market assumes the role of the primary source of systematic risk, these liberalizations introduce an exogenous change that allows for the testing of theoretical predictions.

III. OBJECTIVE

The primary aim of this research is to evaluate the effectiveness of the ARIMA model in forecasting returns for the S&P BSE SENSEX. The study also aims to provide valuable guidance to investors, financial analysts, and policymakers. This research strives to empower stakeholders with practical insights for making informed decisions in the financial markets. Ultimately, the findings of this research aim to benefit individuals and entities engaged in stock market analysis and investment strategies.

IV. DATA AND RESEARCH METHODOLOGY

In the ARIMA, also known as the Box-Jenkins Approach, four stages are sequentially pursued: identification, estimation, diagnostic checks, and forecasting. In this investigation, the data have been gathered from the official BSE website, encompassing a time span of 2 years, 11 months and 4 weeks, comprising a total of 744 trading day observations. This study is focused on analysing the closing value of the S&P BSE SENSEX. This time series analysis executes with the help of EViews software version 10.

1. Identification

When constructing an ARIMA model, the first step is to evaluate the stationarity of the data using informal techniques like graphs and correlograms, as well as formal tests like ADF and PP tests. If non-stationarity is detected, data transformation is used to remove underlying trend patterns. Once stationarity is achieved, potential models are identified using ACF and PACF plots. The PACF helps select the AR component, while the ACF guides the selection of the MA component. Model orders can be inferred from values exceeding the confidence band on the plot. However, it is important to choose a parsimonious model to avoid unnecessary complexity.

2. Estimation

To identify potential ARIMA model candidates, evaluate six key criteria to determine the most suitable model: significant coefficients, SIGMASQ, adjusted R^2 , Akaike Information Criterion (AIC), Schwartz Information Criterion (SIC), and Hannan-Quinn Criterion (HQC). These considerations guide us in selecting the most appropriate ARIMA model.

3. Diagnostic Check

In the diagnostic phase of the Box-Jenkins Method, focus lies on three crucial aspects:

Firstly, ensure the absence of autocorrelation in the residuals of the chosen model. This is accomplished by examining the Ljung-Box Q-statistic.

Secondly, it is of utmost importance to check the stationarity of the residuals in the time series regression. Non-stationary residuals imply an unreliable model and the potential for misinterpretation of results.

Thirdly, to ascertain the stability of the ARIMA model, consider two things:

- a) Verify if the estimated model exhibits covariance stationarity, which is indicated by the inverse AR roots residing within the unit circle.
- b) Ensuring that the estimated process is invertible by ensuring that the inverse MA roots lie inside the unit circle.

If these diagnostic assumptions are not met then must seek a more appropriate model by engaging in overfitting. Overfitting involves adding parameters to the AR or MA components of the model.

4. Forecasting

With the completion of model diagnostics and the subsequent confirmation of the model, the appropriate course of action is now to employ it for the purpose of forecasting.

V. RESULT AND ANALYSIS

1. Identification

Initially plotting the data, it is observed that the time series plot exhibits an overall positive trend (Figure 1).

Figure 1: Graphical representation of close value of S&P BSE SENSEX



Source: Author's own computation

Also, it can be observed from Figure 2 that the plot of ACF exhibits a gradual and linear decay, thereby indicating the non-stationarity of the series at level.

Figure 2: Correlogram of close value of S&P BSE SENSEX

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.988	0.988	729.88	0.000	
2	0.977	-0.012	1443.6	0.000	
3	0.966	0.025	2142.3	0.000	
4	0.956	0.026	2827.3	0.000	
5	0.945	-0.009	3498.4	0.000	
6	0.935	-0.005	4155.8	0.000	
7	0.925	0.001	4799.7	0.000	
8	0.915	-0.004	5430.5	0.000	
9	0.904	-0.008	6047.9	0.000	
10	0.893	-0.053	6650.8	0.000	
11	0.882	0.023	7240.1	0.000	
12	0.872	0.012	7816.6	0.000	
13	0.862	-0.000	8380.6	0.000	
14	0.852	0.009	8932.5	0.000	
15	0.842	-0.001	9472.6	0.000	
16	0.833	0.002	10001.	0.000	
17	0.823	-0.025	10518.	0.000	
18	0.813	-0.006	11023.	0.000	
19	0.802	-0.040	11514.	0.000	
20	0.791	-0.009	11994.	0.000	

Source: Author's own computation

In light of this, the unit root tests are performed. The hypothesis for testing the stationarity of S&P BSE SENSEX series using the ADF and PP tests can be stated as follows:

H₀: The null hypothesis in a unit root test assumes that the time series has a unit root, indicating it is not stationary.

H₁: The alternative hypothesis suggests that the time series does not have a unit root, implying it is stationary.

Table 1: Stationary of the data set

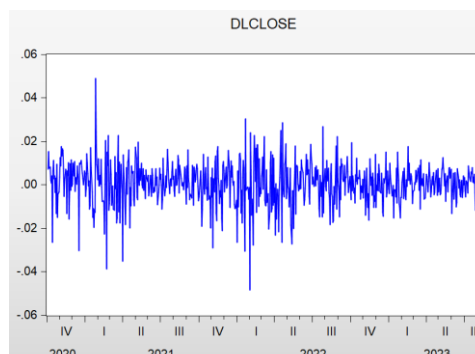
Unit Root Test at Level		
Test Equation	ADF (p-value)	PP (p-value)
None	0.9900	0.9885
Intercept	0.0363	0.0385
Trend and Intercept	0.0485	0.0473

Source: Author's own computation

Based on the information presented in table 1, it can be observed that the p-value exceeds the critical threshold of 0.05 for both tests conducted in the test equation labelled as "None". Consequently, the null hypothesis is deemed acceptable, thus signifying that the data exhibits non-stationarity at the given level. In order to address this non-stationarity concern, it is recommended to initially transform the series into logarithmic values and subsequently apply differencing (DLCLOSE). Henceforth, it is appropriate to proceed with the estimation of a model.

Figure 3 indicates that the series achieves weak stationarity.

Figure 3: Graphical representation of DLCLOSE of S&P BSE SENSEX Returns



Source: Author's own computation

The ADF test and the PP test are employed in order to ascertain the stationarity of this log-differentiated series.

Table 2: Stationary of the data set

Unit Root Test at 1 st Difference		
Test Equation	ADF (p-value)	PP (p-value)
None	0.0000	0.0000
Intercept	0.0000	0.0000
Trend and Intercept	0.0000	0.0000

Source: Author's own computation

Now let's examine the correlogram of the log-differentiated series in order to ascertain the values of p and q for possible models. In this regard, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) provide insights into the potential models.

Figure 2: Correlogram of close value of log-differentiated S&P BSE SENSEX

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.066	0.066	3.2794	0.070
		2 -0.048	-0.053	5.0236	0.081
		3 -0.067	-0.061	8.3889	0.039
		4 0.049	0.056	10.177	0.038
		5 0.044	0.032	11.655	0.040
		6 -0.021	-0.026	11.998	0.062
		7 -0.027	-0.014	12.538	0.084
		8 -0.001	0.002	12.539	0.129
		9 0.005	-0.004	12.556	0.184
		10 -0.055	-0.057	14.809	0.139
		11 -0.012	-0.001	14.922	0.186
		12 -0.005	-0.008	14.944	0.245
		13 -0.019	-0.027	15.206	0.295
		14 -0.019	-0.012	15.476	0.346
		15 -0.001	0.004	15.477	0.418

Source: Author's own computation

After observing the confidence band on the sides. The values that surpass the band indicate a plausible sequence. From the above correlogram, it finds the below parsimony models:

p= (1,2,3)

d= (1)

q= (1,3)

So, Possible Models are:

ARIMA= (1,1,1), (2,1,1), (3,1,1), (1,1,3), (2,1,3) and, (3,1,3).

2. Estimation

In the process of selecting the most suitable ARIMA model from a set of candidates, significant coefficients with preferable p-values of less than 0.05 for both AR and MA terms are sought, ensuring that the included variables exhibit statistical significance. Additionally, lower SIGMASQ is desired, indicating a preference for models with lower volatility. The goal is to maximize the Adjusted R², indicating a better fit of the model to the data. Furthermore, the AIC, SIC, and HQC are minimized to identify the most appropriate model. These criteria collectively guide the selection of the ARIMA model that best suits the analysis.

Table 3: Evolution of the Best Fit Model

Criteria	Model						Best fit Model
	ARMA (1,1)	ARMA (2,1)	ARMA (3,1)	ARMA (1,3)	ARMA (2,3)	ARMA (3,3)	
AR p-value	0.7270	0.1930	0.0314	0.0186	0.2198	0.5704	(3,1) and
MA p-value	0.6092	0.0202	0.0104	0.0213	0.0344	0.4473	(1,3)
SIGMASQ	-	-	8.95479	8.95646	-	-	(3,1)
Adj. R ²	-	-	0.005387	0.005202	-	-	(3,1)
AIC	-	-	-6.472067	-6.471880	-	-	(3,1)
SIC	-	-	-6.447245	-6.447058	-	-	(3,1)
HQC	-	-	-6.462498	-6.462311	-	-	(3,1)

Source: Author's own computation

The ARIMA (3,1,1) model has been identified as the optimal choice, with further details provided in Table 4.

Table 4: ARIMA (3,1,1) model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000714	0.000367	1.943733	0.0523
AR(3)	-0.067677	0.031389	-2.156476	0.0314
MA(1)	0.073156	0.028486	2.568113	0.0104
SIGMASQ	8.95479	3.178536	28.17270	0.0000
R-squared	0.009408	Mean dependent var		0.000715
Adjusted R-squared	0.005387	S.D. dependent var		0.009514
S.E. of regression	0.009489	Akaike info criterion		-6.472067
Sum squared resid	0.066534	Schwarz criterion		-6.447245
Log likelihood	2408.373	Hannan-Quinn criter.		-6.462498
F-statistic	2.339633	Durbin-Watson stat		2.004782
Prob(F-statistic)	0.072201			

Source: Author's own computation

The generalize ARIMA (3,1,1) model written as,

$$\Delta Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} - a U_{t-1} + U_t \quad 1$$

- ΔY_t is the value of the differentiated series at time t .
- c is a constant (intercept).
- ϕ_1, ϕ_2 and ϕ_3 are the autoregressive coefficients corresponding to the lagged values Y_{t-1}, Y_{t-2} and Y_{t-3} respectively.
- a is the moving average (MA) coefficient.
- U_t is a white noise error term at time t .

Applying the ARIMA (3,1,1) coefficients from Table 4, the model takes the following form:

$$DLCLOSE_t = 0.000714 - 0.067677 Y_{t-1} - 0.073156 U_{t-1} \quad 2$$

3. Diagnostic Check

After selecting the best model, it is important to ensure it meets criteria for accurate forecasting. Two important elements come into focus during the diagnostic phase of the Box-Jenkins Method:

I. Absence of Autocorrelation

To validate that the model's residuals exhibit the characteristics of white noise, free from any discernible patterns of autocorrelation, this is tested with the help of Ljung-Box Q-statistic.

H₀: The data demonstrate independent distribution.

H₁: The data do not display independent distribution.

Figure 5: Correlogram of residuals

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.003	-0.003	0.0061	
		2 -0.041	-0.041	1.2541	
		3 -0.002	-0.002	1.2567	0.262
		4 0.049	0.048	3.0743	0.215
		5 0.040	0.040	4.2605	0.235
		6 -0.027	-0.023	4.8015	0.308
		7 -0.026	-0.023	5.2998	0.380
		8 0.003	-0.002	5.3054	0.505
		9 0.007	0.001	5.3412	0.618
		10 -0.058	-0.057	7.8811	0.445
		11 -0.009	-0.005	7.9452	0.540
		12 -0.003	-0.006	7.9517	0.634
		13 -0.021	-0.023	8.2777	0.688
		14 -0.015	-0.011	8.4400	0.750
		15 0.003	0.006	8.4449	0.813

Source: Author's own computation

Figure 5 shows a correlogram of residuals, with small values within the 95% confidence interval. This indicates that the residuals are significant and supports the null hypothesis. The residuals are independent, indicating no autocorrelation. This supports the use of the ARIMA (3,1,1) model and suggests the error terms exhibit white noise characteristics.

II. Stationary Check of the Residuals

Evaluating the stationarity of residuals in a time series regression model is crucial for confirming the model's credibility and the dependability of its results. When residuals exhibit non-stationarity, it indicates potential issues with the model's validity and raises the risk of misinterpreting the findings.

H₀: The null hypothesis in a unit root test suggests the presence of a unit root in the time series, indicating non-stationarity.

H₁: The alternative hypothesis states that there is no unit root in the time series, indicating stationarity.

Table 5: Stationary of the Residuals

Unit Root Test at Level		
Test Equation	ADF (p-value)	PP (p-value)
None	0.0000	0.0000
Intercept	0.0000	0.0000
Trend and Intercept	0.0000	0.0000

Source: Author's own computation

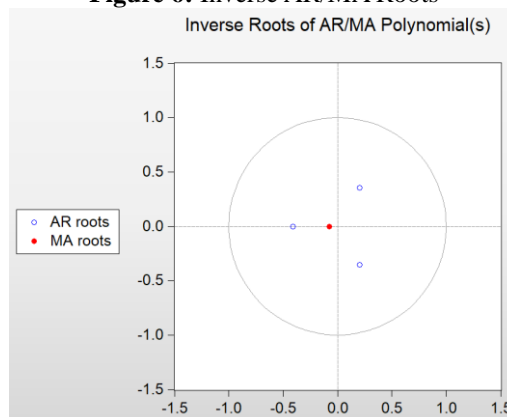
Table 5 shows that all values are lower than the significance level of 0.05, resulting in the rejection of H₀. Consequently, all residuals exhibit stationarity at the level, ensuring the model's validity, forecast reliability, and statistical inference accuracy.

III. Stability Check

In assessing the stability condition within the ARIMA model, two critical aspects are examined:

- Verification of covariance stationarity: The roots of the inverse AR components should reside within the unit circle.
- Confirmation of invertibility: The roots of the inverse MA components should also remain inside the unit circle.

Figure 6: Inverse AR/MA Roots



Source: Author's own computation

As depicted in the above figure, it is evident that all the inverse roots are contained within the unit circle. The ARIMA (3,1,1) model fulfills the stability conditions, and the error terms exhibit characteristics of white noise. This positions the analysis in a favorable position for forecasting future S&P BSE SENSEX returns values.

4. Forecasting

I) Static Forecasting within the Sample

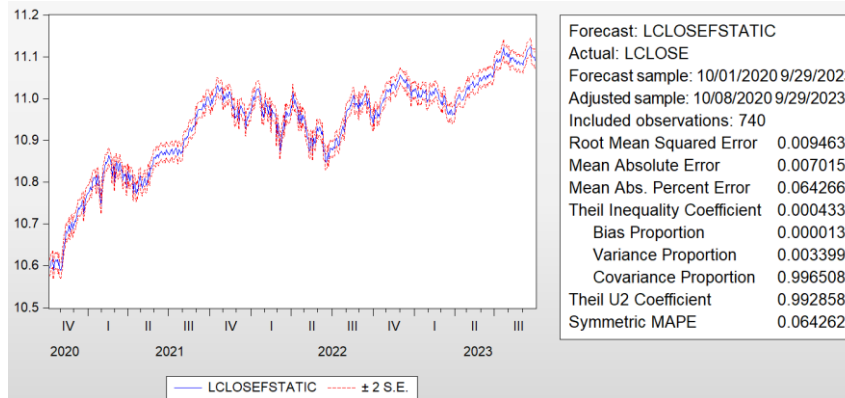
The ARIMA (3,1,1) model has been used to forecast the closing price returns of S&P BSE SENSEX spanning from October 01, 2020, to September 29, 2023. Figure 7 shows the forecast and actual values, with a confidence interval. The forecast performance metrics are also provided. The model is of high quality, with a remarkably low Theil coefficient (0.000433). Figure 7 also shows a strong alignment between actual and forecasted values.

An ideal forecast should be unbiased, accurate, and free from random fluctuations. The bias proportion value (0.000013) in Figure 7 indicates a positive outcome, suggesting highly accurate forecasts without systematic bias.

The majority of the variability in the time series data is accounted for by the model's forecasts, as indicated by the variance proportion (0.0003399) in Figure 7. The covariance proportion (0.996508) is also high, reflecting satisfactory performance.

With satisfactory bias and variance proportion values, the model is suitable for forecasting. The Mean Absolute Percentage Error (MAPE) of 0.06% indicates a level of accuracy commonly considered acceptable in practical scenarios.

Figure 7: Actual and Forecast of the S&P BSE SENSEX form 10/01/2020 to 9/29/2023

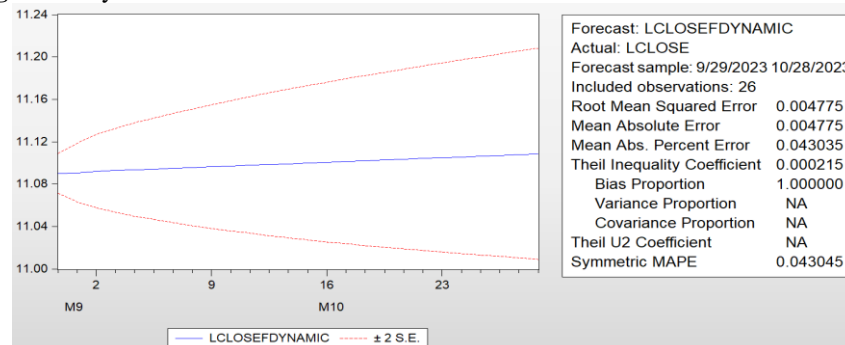


Source: Author's own computation

II) Dynamic Forecasting Out of the Sample

The process of dynamic forecasting begins by determining the period for future predictions. A 29-day horizon was chosen for this analysis. The ARIMA (3,1,1) model is used to generate the forecasts, along with 95% confidence intervals. Figure 8 visually represents the dynamic forecasts by overlaying actual historical values with forecasted values. The shaded region within the plot represents the 95% confidence interval. This approach allows for assessing the model's performance and understanding forecasted trends and variations. The data and forecast information from September 29, 2023 to October 28, 2023 are presented in ANNEXURE A1, utilising the ARIMA model (3,1,1).

Figure 8: Dynamic Forecast of the S&P BSE SENSEX form 9/29/2023 to 10/28/2023



Source: Author own computation

The analysis began with a 29-day horizon for future predictions using an ARIMA model. It included 95% confidence intervals to account for uncertainty. Figure 8 shows a visual representation of the forecasts, with the shaded region indicating the confidence interval.

The RMSE and MAE metrics indicate the accuracy of the forecasts, with minimal errors. The MAPE shows a relative error of 0.043%, demonstrating precision. The Theil Inequality Coefficient is impressively low, indicating high-quality performance.

VI. CONCLUSION

For stock market forecasting, the ARIMA methodology, also known as Box-Jenkins, is widely employed. It enables traders, investors, portfolio managers, and financial institutions to build robust financial models, effectively manage risks, and make well-informed decisions. For forecasting, a robust ARIMA (3,1,1) model was adopted. This study focuses on forecasting S&P BSE SENSEX returns using ARIMA. The diagnostics, stability, and forecasting skills of the model were assessed and consistently produced positive results. While ARIMA does not guarantee profits, it does provide useful information for decision-making. Other methodologies, like as GARCH models, can be useful for volatility modelling and forecasting.

REFERENCES

1. Chari, A., & Henry, P. B. (2004). Risk sharing and asset prices: Evidence from a natural experiment. *The Journal of Finance*, 59(3), 1295–1324. Available at: <https://doi.org/10.1111/j.1540-6261.2004.00663.x>.

2. Gupta, S. K., & Rao, R. P. (2018). The causal relationship between financial development and economic growth: An experience with BRICS economies. SpringerLink. *J Soc Econ Dev*, 20(2), 308–326. Available at: <https://doi.org/10.1007/s40847-018-0071-5>.
3. Jiahan, L., & Ilias, T. (2017). Equity premium prediction: The role of economic and statistical constraints. *Journal of Financial Markets*, 36(C), 56–75. Available at: <https://doi.org/10.1016/j.finmar.2016.09.001>.
4. Kaufman, G. G. (1995). Comment on systemic risk. *Research in Financial Services: Banking, Financial Markets, and Systemic Risk*, 7, 47–52.
5. Kim, J., Lim, K., & Shamsuddin, A. (2011). Stock return predictability and the adaptive markets hypothesis: Evidence from century-long U.S. data. *Journal of Empirical Finance*, 18(5), 868–879. Available at: <https://doi.org/10.1016/j.jempfin.2011.08.002>.
6. Mallikarjuna, M., & Rao, R. P. (2019). Evaluation of forecasting methods from selected stock market returns. *Finance Innov*, 5(40). Available at: <https://doi.org/10.1186/s40854-019-0157-x>.
7. Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science*, 60(7), 1772–1791. Available at: <https://doi.org/10.1287/mnsc.2013.1838>.
8. Zhu, X., & Zhu, J. (2013). Predicting stock returns: A regime-switching combination approach and economic links. *Journal of Banking & Finance*, 37(11), 4120–4133. Available at: <https://doi.org/10.1016/j.jbankfin.2013.07.016>.

ANNEXURE

A1: Dynamic Forecasted value of S&P BSE SENSEX

Date	Close Forecast
9/29/2023	11.09003
9/30/2023	11.09062
10/02/2023	11.09201
10/03/2023	11.09276
10/04/2023	11.09349
10/05/2023	11.09415
10/06/2023	11.09486
10/07/2023	11.09558
10/09/2023	11.0963
10/10/2023	11.09701
10/11/2023	11.09772
10/12/2023	11.09844
10/13/2023	11.09915
10/14/2023	11.09986
10/16/2023	11.10058
10/17/2023	11.10129
10/18/2023	11.10201
10/19/2023	11.10272
10/20/2023	11.10343
10/21/2023	11.10415
10/23/2023	11.10486
10/24/2023	11.10558
10/25/2023	11.10629
10/26/2023	11.107
10/27/2023	11.10772
10/28/2023	11.10843