Proceedings of CONF-BPS 2026 Symposium: GenAI, Labour Markets, and the Economics of Human and Financial Capital
DOI: 10.54254/2754-1169/2026.LD31587

Study of the ARIMA Model in Financial Time Series Forecasting

Jiayi Wang

Faculty of Economics, Sophia University, Tokyo, Japan
oukaeki35@gmail.com

Abstract. As stock market data becomes more transparent and accessible to the public, along with the application of advanced financial technologies, the demand for accurate financial market forecasts is rising. This growing need is essential for developing effective investment strategies. There are many econometric models applied to financial time series data forecasts, each based on its unique assumptions and subjected to different situations. This paper discusses the ARIMA model in econometrics for time series data forecasting, specifically its application in predicting Netflix's closing stock prices. It analyzes the methodology used in this case study, highlighting the limitations of the ARIMA model in dynamic forecasting and its challenges in modeling financial time series data. After replicating steps in the Netflix stock price forecast, it finds that although the ARIMA model performs well in the short term, or the one-step-ahead forecast, it lacks explanatory power in dynamic forecasts, especially in the case of dealing with financial time series data.

Keywords: ARIMA, Time Series Analysis, Stock market, Econometrics

1. Introduction

The ARIMA model has been widely employed in financial time series forecasting due to its functionality in handling non-stationary data and capturing the short-run dynamics. Since the published work of Box and Jenkins, ARIMA has become a fundamental tool in econometrics in terms of forecasting macroeconomic as well as financial variables [1]. Several historical studies have illustrated the effectiveness of the ARIMA model in short-term financial forecasting. For example, Par and Lin applied a hybrid ARIMA and support vector machines (SVMs) to forecast the Nasdaq index and concluded their combined effectiveness in modeling the short-term trends [2]. Similarly, Atsalakis and Valavanis applied multiple techniques, including the Artificial Neural Networks (ANNs), Genetic Algorithms, and the ARIMA model, in the stock market prediction and confirmed the forecasting strength of the ARIMA [3]. However, ARIMA has limitations, especially in dealing with nonlinearities and volatility clustering, which are often observed in the financial time series data [4]. To overcome the issue, researchers often adopt hybrid models combining ARIMA with ANNS, GARCH, and Machine learning models [5]. Thus, recognizing ARIMA's inherent constraints is critical. Khan and Alghulaiakh applied ARIMA to forecast Netflix's stock prices and assessed its financial time series predictive efficacy [6]. By replicating their work, this paper clarifies ARIMA's mechanism and drawbacks in time series analysis. This not only deepens

DOI: 10.54254/2754-1169/2026.LD31587

understanding of ARIMA's practical boundaries but also provides a grounded reference for selecting appropriate models in financial forecasting.

2. Time series analysis

A defining feature of time series data is its collection over time intervals, demonstrating how variables evolve. This characteristic contravenes the i.i.d. assumption. Verbeek, in "A Guide to Modern Econometrics," defines time series observations as realizations of random variables within stochastic processes. In economics, variables like output, interest rates, and inflation vary according to business cycles [7]. Similarly, in financial markets, security prices reflect economic conditions, shocks, and investor sentiment, indicating that time series data often follows stochastic trends, challenging the independence of observations and the uniformity of their probability distributions. Statistically, the OLS estimator fails to achieve the Best Linear Unbiased Estimator (BLUE) properties when applied to time series data due to violations of the assumptions of independence, identically distributed (i.i.d) data, and homoscedasticity. Autocorrelation in time series causes current values to depend on their past values, while the variance of error terms is not consistent, affecting the reliability of test statistics. This can result in a higher likelihood of Type I and Type II errors [8].

3. Methodology

3.1. ARIMA model

The Auto Regressive Integrated Moving Average (ARIMA) model is a widely used method for forecasting future time series values, consisting of three main components: the Auto Regressive model, the integration of data, and the Moving Average model. It captures the influence of lagged values on current observations and is especially relevant in financial markets, where it reflects investors' expectations for future stock prices based on current values. The Moving Average aspect effectively accounts for short-term market dynamics by considering past errors. Additionally, data must be stationary—having a constant mean and variance—through appropriate degrees of differencing; predictions are unreliable if this condition is not met.

$$y_t = \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q} \quad (1)$$

3.2. Data description

Khan and Alghulaiakh conducted a five-year analysis of Netflix's daily stock data to predict its price dynamics over the subsequent 100 days, focusing on the adjusted closing price as the most accurate measure of stock value. Due to temporary restrictions on download permissions from Yahoo! Finance during the research, the replication utilized data sourced from MarketWatch, which did not provide the adjusted closing price. Therefore, the differences in available data likely contributes to differences between the findings in the original article and this replication. Specifically, the original study used closing prices, while this replication emphasizes adjusted closing prices. Additionally, variations in how daily trading volume is calculated between Yahoo! Finance and MarketWatch could also affect statistical outcomes. Unlike the targeted article, which employed R for model implementation, this paper utilized EViews 13 for analysis, with all figures generated through the author's work in EViews.

DOI: 10.54254/2754-1169/2026.LD31587

3.3. Visualizing Netflix's closing price to identify trends, seasonality, and patterns

Following the data preprocessing procedure, three additional time series datasets have been created based on the original value of the closing price. Specifically, W represents the weekly average of Netflix's closing price, M indicates the monthly average of Netflix's closing price, and Y denotes the yearly average of Netflix's closing price. The additional datasets are generated using the moving average method, aiming to reduce the noise of the original dataset to reveal the general pattern of the closing price. During the implementation, the weekly average is calculated by assuming n = 5, the monthly average is calculated by assuming n = 20, and the yearly average is calculated by assuming n = 252, in which n represents the trading days.

As illustrated in Figure 1, seasonality and trending can be detected. The closing price fluctuates around different averages based on the different seasons each year can be direct evidence showing seasonality, which leads to the fact that the data is not stationary. In other words, the mean of the closing price varies throughout the entire period, and its variance also shows an inconsistent form.

Figure 1. Historical closing price of Netflix

3.4. Augmented dickey-fuller test on the closing price of Netflix after differencing

Table 1. ADF test D(CLOSE)

Null Hypothesis	D(CLOSE) has a unit root	
Exogenous	Constant, Linear Trend	
Lag Length: 0 (Automatic - based on SIC, maxlag=10)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-38.21242	0.0000
Test critical values	1% level	-3.965352
	5% level	-3.413385
	10% level	-3.128727
*MacKinnon (1996) one-sided p-values.		

After taking the first difference of the closing price of Netflix, the ADF test statistic in Table 1 reveals that the data is stationary, as the ADF test statistic is significant at all levels. However, the first difference in the closing price here does not generate a meaningful interpretation. To make the data more interpretable, the log difference, representing the percentage change of the closing price, will be implemented in the later ARIMA model construction.

DOI: 10.54254/2754-1169/2026.LD31587

3.5. ARIMA model selection

Table 2. Correlogram

Date	07/07/25	Time	22:12	
Sample (adjusted)		4/08/2015 4/07/2020		
Included observations		1260 after adjustments		
	AC	PAC	Q-Sat	Prob
1	-0.028	-0.028	1.0258	0.311
2	0.032	0.031	2.3053	0.316
3	0.032	0.034	3.6583	0.303
4	-0.013	0.012	3.8566	0.426
5	-0.002	-0.075	10.429	0.064
6	-0.015	0.020	10.720	0.097
7	-0.014	0.019	10.924	0.140
8	0.067	0.060	16.602	0.034
9	0.029	0.023	17.670	0.039
10	-0.036	-0.038	19.359	0.036

Table 2 shows the correlogram of the closing price after taking the log difference. To determine how many lags to include in ARMA(p,q), we are interested in observing where the Autocorrelation and the Partial Autocorrelation tails off. Observed from the table, we consider two tentative ARIMA specifications for further testing: ARIMA(5,1,5) and ARIMA(5,1,8), where lag 5 and lag 8 in ACF show marginal significance. However, due to the limited clarity in cutoff patterns and the risk of overfitting the data with high lag orders, information criteria (AIC/BIC) and residual diagnostics will be implemented in the subsequent steps to finalize the model selection.

4. Results

By comparing Tables 3 and 4 of the two tentative ARIMA models, evidences show that the ARIMA(5,1,8) outperforms the ARIMA(5,1,5). First of all, the ARIMA(5,1,8) has a higher R-squared value (0.009417) than the ARIMA(5,1,5) model. Although a higher R-squared value in the ARIMA model might indicate overfitting the data or a lack of parsimony, this problem is less of a concern in our case since both ARIMA models include the same number of lags. In addition, the ARIMA(5,1,8) model has a smaller Akaike info criterion (-4.425) and the Schwarz criterion (-4.418), which means that compared with the ARIMA(5,1,5), the ARIMA(5,1,8) better fits the data while maintaining the parsimony. Furthermore, the ARIMA(5,1,8) has a smaller sigma-squared value, which indicates that the ARIMA(5,1,8) is less volatile compared to the ARIMA(5,1,5), and the prediction error is smaller. Combining all the information from the summary table, we can rationally choose the ARIMA(5,1,8) as the model of best fit.

Table 3. Summary table of ARIMA (5, 1, 5)

Dependent Variable:D(LOG(CLOSE))Method:ARMA Maximum Likelihood (OPG - BHHH)Date:07/07/25 Time:22:51Sample:4/08/2015 4/07/2020Included observations:1260Convergence achieved after 11 iterationsCoefficient covariance computed using outer product of gradients

DOI: 10.54254/2754-1169/2026.LD31587

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001421	0.000655	2.170221	0.0302
AR(5)	-0.071114	0.023946	-2.969717	0.0030
MA(8)	-0.063811	0.027226	-2.343731	0.0192
SIGMASQ	0.000697	1.38E-05	50.37447	0.0000
R-squared	0.009417	Mean dependent var		0.001442
Adjusted R-squared	0.007451	S.D. dependent var		0.026529
S.E. of regression	0.026035	Akaike info criterion		-4.420349
Sum squared resid	0.877723	Schwarz criterion		-4.408706
Log likelihood	2791.762	Hannan-Quinn criter.		-4.418889
F-statistic	3.980268	Durbin-Watson stat		2.054489
Prob(F-statistic)	0.007773			
Inverted AR Roots	.80	.25-.76i	.25+.76i	-.65-.47i
	-.65+.47i			
Inverted MA Roots	.83	.26-.79i	.26+.79i	-.67+.49i
	-.67-.49i			

Table 4. Summary table of ARIMA(5, 1, 8)

Dependent Variable:D(LOG(CLOSE))Method:ARMA Maximum Likelihood (OPG - BHHH)Date:07/07/25 Time:22:52Sample:4/08/2015 4/07/2020Included observations:1260Convergence achieved after 11 iterationsCoefficient covariance computed using outer product of gradients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001421	0.000655	2.170221	0.0302
AR(5)	-0.071114	0.023946	-2.969717	0.0030
MA(8)	-0.063811	0.027226	-2.343731	0.0192
SIGMASQ	0.000697	1.38E-05	50.37447	0.0000
R-squared	0.009417	Mean dependent var		0.001442
Adjusted R-squared	0.007051	S.D. dependent var		0.026529
S.E. of regression	0.026435	Akaike info criterion		-4.425019
Sum squared resid	0.877723	Schwarz criterion		-4.408706
Log likelihood	2791.762	Hannan-Quinn criter.		-4.418889
F-statistic	3.980268	Durbin-Watson stat		2.054489
Prob(F-statistic)	0.007773			
Inverted AR Roots	.48+.35i-.59	.48+.35i	-.18+.56i	-.18-.56i
Inverted MA Roots	.71-.00-.71i	.50-.50i-.50+.50i	.50-.50i-.50+.50i	.00+.71i-.71

Table 5. Residual diagnostic test of ARIMA(5,1,8)

Date:07/08/25 Time:00:16 Sample (adjusted):4/08/2015 4/07/2020 Q-statistic probabilities adjusted for 2 ARMA terms			
AC	PAC	Q-Stat	Prob

DOI: 10.54254/2754-1169/2026.LD31587

1	-0.028	-0.028	1.0135
2	0.032	0.032	2.3382
3	0.031	0.033	3.5588
4	-0.013	-0.013	3.7814
5	-0.002	-0.005	3.7869
6	-0.014	-0.014	4.0376
7	0.016	0.016	4.3558
8	0.000	0.002	4.3560
9	0.026	0.026	5.1892
10	-0.038	-0.039	7.0613

Results inTable 5 indicate that the residuals of the ARIMA(5,1,8) model show no significant autocorrelation within the first 10 lags at a 5% significance level, as evidenced by p-values exceeding this threshold. This supports the conclusion that the model accurately captures the data's structure and that the residuals follow white noise. Additionally, the correlogram demonstrates that ACF and PACF spikes remain within the 95% confidence interval, reinforcing the model's selection for forecasting Netflix's closing prices.

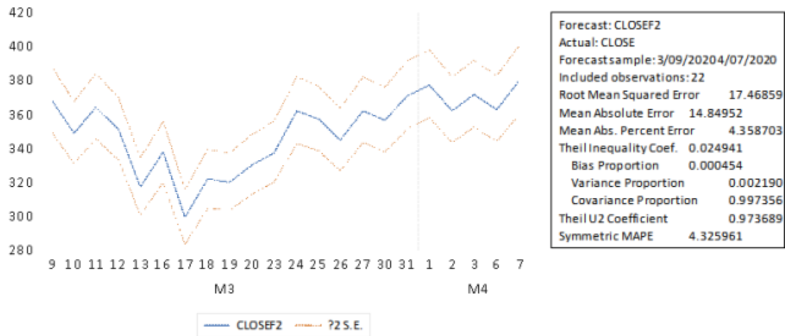

Figure 2. Static forecast of ARIMA(5,1,8)

Figure 2 illustrates the static forecast results of the ARIMA(5,1,8) model for Netflix's closing price. The actual closing prices are shown by a blue line, with dashed lines representing forecasted values and confidence intervals. The model effectively tracks actual data dynamics with minor deviations during sharp movements. Accuracy metrics indicate strong performance, with a Root Mean Squared Error of 17.47 and a Mean Absolute Error of 14.85. A Symmetric MAPE of approximately 4.36% demonstrates high predictive reliability, while a Theil Inequality Coefficient of 0.0249 indicates unsystematic forecast errors. This suggests that ARIMA(5,1,8) is a robust model for forecasting closing price movements.

5. Analysis of the forecasting results and the limitations of the ARIMA model

Comparative analysis of dynamic forecasting results, illustrated in Figures 3 and 4, shows distinct differences in predicted closing price movements. This paper indicates an upward trend in closing prices over 100 days post-April 7, 2020, while the referenced article suggests a stable trend with slight declines for Netflix. These discrepancies may stem from varying data usage, with this study utilizing standard closing prices versus the adjusted prices in the article, differences in ARIMA model methodologies, and potential axis scaling issues, as the targeted article employs normalized

DOI: 10.54254/2754-1169/2026.LD31587

scales. Notably, both forecasts reveal poor, insignificant results, as they project a linear trend with diverging confidence intervals, rendering the predictions impractical for real-world application.

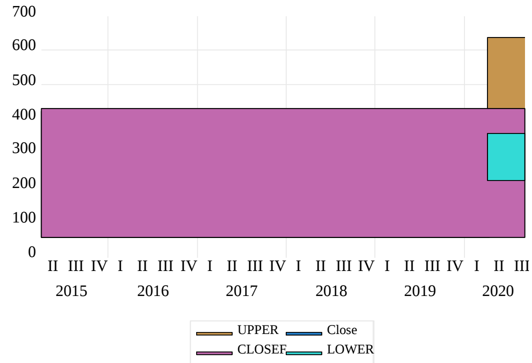

Figure 3. Dynamic forecast of ARIMA(5,1,8)

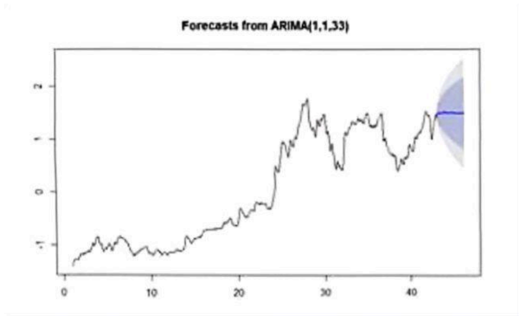

Figure 4. Dynamic forecast of ARIMA(1,1,33)

6. Conclusion

The replication results highlight the limitations of the ARIMA model in out-of-sample forecasting and its ability to capture financial time series dynamics. While ARIMA can track overall trends in initial forecasts, its performance declines over time, marked by widening confidence intervals and growing discrepancies from actual values. This deterioration stems from the model's reliance on past forecasted values rather than actual data, which can lead to substantial deviations due to small model errors or misspecifications. Additionally, ARIMA's linear nature and its assumption of stationary processes stand in contrast to the characteristics of financial markets, which often display volatility clustering, regime shifts, and non-linear relationships. As a result, ARIMA may be adequate for short-term forecasts under stable conditions but typically underperforms in dynamic, volatile environments. This underscores the necessity for integrating ARIMA with more advanced modeling techniques, such as GARCH or machine learning approaches, for enhanced forecasting performance in complex financial contexts.

References

[1] Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day.

[2] Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 33(6), 497–505. https: //doi.org/10.1016/j.omega.2004.07.024

[3] Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. Expert Systems with Applications, 36(3), 5932–5941. https: //doi.org/10.1016/j.eswa.2008.07.006

DOI: 10.54254/2754-1169/2026.LD31587

[4] Petrică, A.-C., Stancu, S., & Tindeche, A. (2016). Limitation of ARIMA models in financial and monetary economics. Theoretical and Applied Economics, 23(4(609)), 19–42.
[5] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175. https: //doi.org/10.1016/S0925-2312(01)00702-0
[6] Khan, S., &Alghulaiakh, H.(2020). ARIMA model for accurate time series stocks forecasting. International Journal of Advanced Computer Science and Application (IJACSA), 11(7).
[7] Verbeek, M. (2017). A guide to modern econometrics (5th ed.). Wiley.
[8] Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert Systems with Applications, 38(5), 5311–5319.