2025 SICASH MULTICONFERENCE

International Conference on Actuarial Sciences (ICActS)
Available on-line at: **www.sicash.sliit.lk**

Predictive Model for the SPDR S&P 500 ETF (SPY) using Volatility Analysis Approach

N. I. Musharraff[1]*, W. S. C. Fernando[1], T. R. Godage[1], J. M. T. S. Jayasooriya[1], H. A. A. T. P. Siriwardhana[1], T. A. Samasundara[1], M. L. Guruge[1], T. S. G. Peiris[1]

[1]*Department of Mathematics and Statistics, Faculty of Humanities and Sciences, Sri Lanka Institute of Information Technology (SLIIT), Malabe, 10115, Sri Lanka*

Corresponding author*: hs23367180@my.sliit.lk

Abstract

The S&P 500 (Standard & poor's 500) is one of the most widely followed equity indices in the world. The SPDR S&P 500 ETF Trust (SPY) is used to track the performance of the S&P 500 index as closely as possible and can also be traded in the stock exchanges. Not many studies have been carried out to forecast daily closing prices of SPY for recent years. This study presents a time series analysis and forecasting of the daily closing prices of the SPY index. The dataset extends from 2000 to 2025, capturing key financial events, market movements and long-term growth trends. Due to high volatility, we were forced to consider variance equation in additional to the mean equation and the best fitted model identifies is ARIMA (1,1,1) + GARCH (1,1). The best fitted model was selected by comparing multiple parsimonious models using AIC, HQC and Schwarz Criterion. Diagnostic checks confirmed that the residuals were white noise, indicating that the model sufficiently captured the fundamental structure of the data. The model was trained using the historical data from 3rd January 2000 to 2nd April 2025 and validated for 3rd April 2025 to 13th June 2025. The percentage error for the testing dataset had a range of -13.9% to 3.9%. The intention was to build a model and analyse the behaviour of SPY and provide short-term forecasts that could assist investors, business analysts, and policymakers in decision-making. The results are a testament to the high predictive ability of the postulated model, which also provides a methodological framework for future predictions.

Keywords: S&P 500, SPY, ARIMA, GARCH, AIC

Introduction

The ability to accurately forecast the financial markets has long been a focal point in the fields of Finance, Data science and Economics due to the profound role it plays in both risk management and investment decision making. Among the various financial instruments, the S&P 500 index is one that stands out, due to its prominence in manifesting the broad trend of the U.S. equity markets (Nickolas, 2024). Hence the SPDR S&P 500 ETF Trust (SPY), which tracks the performance of that very index and which is also one of the most actively traded and highly influential exchange-traded funds (ETFs) is of paramount importance in both acting as a proxy to the U.S. Equity markets and also in reflecting investor sentiment and macroeconomic trends.

The SPY is one of the most profound financial instruments in U.S. capital markets offering current and prospective investors a holistic exposure to benchmark the performance of the Equity markets via a single tradable security. SPY represents the collective movement of 500 of the largest U.S. companies

https://doi.org/10.54389/DZQE4018

spanning various sectors and launched in 1993 (State Street global advisors, 2025). Its high liquidity, low transaction fees and high transparency make it an attractive investment vehicle for both institutional and retail investors. Beyond acting as a financial instrument, SPY plays a pivotal role in financial research, particularly in forecasting long term trends, volatility and market behavior due to its consistent historical data and real time price action.

Several models have been applied to forecast financial time series such as ARIMA, ARCH and GARCH, there exists significant debate regarding the accuracy and predictive power of these models, especially under conditions of high market volatility and structural shifts. Furthermore, most of the research has been centered around the S&P 500 index itself and no research has been done with regard to SPY despite the differences in dividend payout structure, liquidity and trading behavior. Overfitting risks in model performance highlights the need to postulate a more updated model for daily price movements, validate the model and use the model to predict future price movements.

Materials and Methods

Secondary Data

The daily SPDR S&P 500 ETF Trust (SPY) prices in US dollars from 3rd January 2000 to 13th June 2025 were obtained. This Data set excludes weekends and holidays. The dataset from 3rd January 2000 to 2nd April 2025 was used to train the model and the data ranging from 3rd April 2025 to 13th June 2025 was used to validate the model. Statistical software packages such as Minitab and E-views 12 were used to perform this analysis.

Methodology

An autoregressive Moving Average (ARMA) model is a popular time series forecasting method that combines two key components, the Autoregressive (AR) part and the Moving Average (MA) part. (Box & Jenkins, 1970). The AR model with order p and MA model with q are shown in equation (1) and equation (2) respectively. $\{e_t\}$ is purely random process with zero mean and constant variance. Thus ARMA(p,q) model can be expressed as shown in (3).

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \ldots + \emptyset_p Y_{t-p} + \varepsilon_t \quad \text{---------------(1)}$$

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \quad \text{---------------(2)}$$

$$Y_t = \emptyset_1 Y_{t-1} + \ldots + \emptyset_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \quad \text{---------------(3)}$$

Where, f_i is the Autoregressive coefficients and q_i is the Moving Average coefficients. The main assumption in the ARMA model is, it requires stationarity. (Mean, variance, autocovariance constant over Time). If a series performs non-stationarity a technique called Differencing is used to make the series stationary. (Box & Jenkins, 1970). After differencing, the ARMA model can be expressed as ARIMA(p,d,q) where d is the number of differences to achieve stationarity and p, q denotes the order of AR and MA terms respectively.

However, Financial time series such as Exchange rates and stock prices are more volatile because of several economic and structural factors. Thus, the errors are heteroscedastic, which means the variance of error terms is not constant. To model conditional variance, ARCH (Engle, 1982) /GARCH (Bollerslev, 1986) models are used typically to build the variance equation to capture the clustering volatility. Since volatility persisted for a long time in the SPY stock prices, we have used a GARCH model which is better for long-lasting volatility effects.

The model equation for GARCH (p,q) where p is the order of the GARCH terms and q is the order of the ARCH terms) is written as:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots\ldots + \alpha_q u_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \ldots\ldots\ldots + \beta_p \sigma_{t-p}^2 \quad \text{--------------(4)}$$

Generally, GARCH balances model complexity with better performances in most financial time series.

Results and discussion

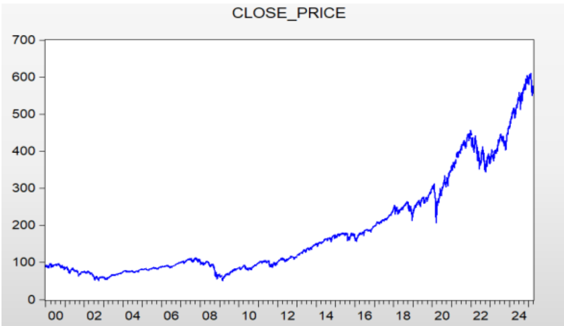

Figure 1: *Temporal Variability of original series*

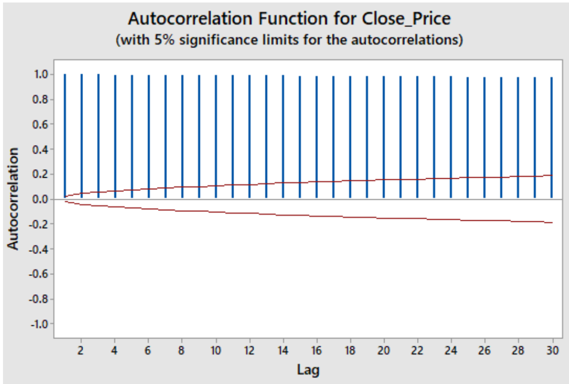

Figure 2: *ACF plot of original series*

Figure 1 shows the temporal variability of daily closing stock prices of SPY from 3rd January 2000 to 2nd April 2025. It varies between $50.3795 and $611.092 with a mean of $185.3521 and standard deviation of $136.737. Jarque-Bera test statistic showed significant. It can be concluded that data is deviated from normality at 5% significant level. ACF is shown in Fig.2. Figure 2 shows that all the autocorrelations are significant at every lag, indicating the original series is not stationary. This was confirmed by the Augmented Dickey-Fuller test (Table1).

Table 1: *ADF Test for original series*

Null Hypothesis: CLOSE_PRICE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 10 (Automatic - based on SIC, maxlag=33)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.635595	0.9765
Test critical values:	1% level	-3.959484	
	5% level	-3.410513	
	10% level	-3.127024	

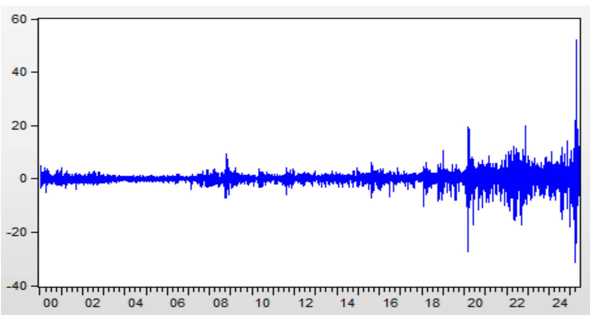

Figure 3: *Time series plot for 1st difference series*

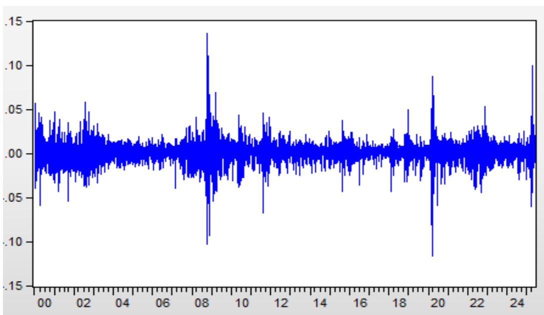

Figure 4: *Time series plot for 1st difference of log*

Table 2: *ADF test for 1^{st} difference of log series*

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-86.77523	0.0001
Test critical values:	1% level	-3.959482	
	5% level	-3.410511	
	10% level	-3.127024	

According to figure 3, the volatility can be seen throughout the period, and it increases towards the later part. On such an occasion the original series can be transformed into its log difference series to reduce Heteroskedastic of variance. This will lead to better stationarity and better forecasting. Figure 4 indicates that 1^{st} difference of log series is better stationarity than the 1^{st} differenced series. The stationarity of the log differenced series was confirmed statistically by the ADF test shown in table 2. Therefore, to identify the possible ARIMA models, ACF and PACF plots for 1^{st} differenced of log series was considered.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.084	-0.084	44.928	0.000
		2	-0.021	-0.028	47.758	0.000
		3	0.004	-0.000	47.863	0.000
		4	-0.021	-0.021	50.645	0.000
		5	-0.012	-0.015	51.536	0.000
		6	-0.035	-0.038	59.181	0.000
		7	0.035	0.028	67.029	0.000
		8	-0.025	-0.022	71.135	0.000
		9	0.031	0.028	77.122	0.000
		10	-0.013	-0.012	78.249	0.000

Figure 5: *ACF and PACF plots of 1^{st} difference log series*

ACF and PACF patterns for the 1^{st} differenced log series are shown in figure 5. According to this theoretical ACF & PACF it suggests that ARIMA (1,1,0), ARIMA (0,1,1), ARIMA (1,1,1) may be some possible models as lag 1 in both ACF & PACF is significant. The errors were checked in each model, and it was found that errors in each model are not randomly distributed. Thus, the ACF and PACF of squared residuals were taken and it showed that there are significant spikes at many lags. Thus, the ARCH effect for mean equations were tested by Heteroskedasticity test for residuals.

Table 3: *Heteroskedasticity Test for ARIMA models*

Model	P value
ARIMA (1,1,0)	< 0.05
ARIMA (0,1,1)	< 0.05
ARIMA (1,1,1)	< 0.05

According to table 3, the p values for test statistics for each mean equation were statistically significant, which concludes that significant ARCH effect exists in both models at 5% significant level. Thus, variance equation for the conditional variance is needed. Hence, the comparison of mean equations with GARCH (1,1) was conducted.

Table 4: *Comparison of three postulated models with volatility*

Model	MA Term	AR Term	ARCH Term	GARCH Term	AIC	SC	HQC
ARIMA (0,1,1)+GARCH (1,1)	Sig	-	Sig	Sig	-6.4600	-6.4547	-6.4582
ARIMA (1,1,0)+GARCH (1,1)	-	Sig	Sig	Sig	-6.4605	-6.4551	-6.4586
ARIMA (1,1,1)+GARCH (1,1)	Sig	Sig	Sig	Sig	-6.4620	-6.4556	-6.4598

According to table 4, ARIMA (1,1,1) - GARCH (1,1) has all significant coefficients and the lowest AIC, SC and HQC values compared to the other models. Therefore ARIMA (1,1,1) - GARCH (1,1) is set to be the best model for forecasting.

Table 5: *Results of Heteroskedasticity test*

Heteroskedasticity Test: ARCH

F-statistic	1.271915	Prob. F(1,6345)	0.2595
Obs*R-squared	1.272061	Prob. Chi-Square(1)	0.2594

Table 5 shows that the p value is greater than 0.05 for Heteroskedasticity test statistic which means that there is no longer a significant ARCH effect. Thus, it can be concluded by 95% of confidence that ARIMA (1,1,1) + GARCH (1,1) model has been able to exert a significant control over the ARCH effect.

Date: 06/30/25 Time: 18:30
Sample (adjusted): 1/06/2000 4/02/2025
Q-statistic probabilities adjusted for 3 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.012	0.012	0.9226	
		2	-0.002	-0.002	0.9464	
		3	0.003	0.003	0.9891	
		4	0.009	0.009	1.5431	0.214
		5	-0.016	-0.016	3.1119	0.211
		6	-0.012	-0.011	3.9638	0.265
		7	0.024	0.024	7.6543	0.105
		8	0.004	0.004	7.7715	0.169
		9	0.005	0.006	7.9559	0.241
		10	0.009	0.009	8.5263	0.288
		11	0.019	0.018	10.782	0.214
		12	0.017	0.017	12.605	0.181
		13	0.014	0.015	13.941	0.176
		14	-0.008	-0.009	14.399	0.212
		15	-0.023	-0.023	17.775	0.123
		16	0.015	0.016	19.301	0.114
		17	0.012	0.012	20.194	0.124

Figure 6: *ACF and PACF of residuals for the fitted model*

According to figure 6, Q statistic in every lag is not significant concluding with 95% confidence that the errors are distributed randomly in the fitted model. The Jarque-Bera test statistic for residuals showed significant(p=0.00) It can be concluded that residuals are deviated from normality at 5% significant level. But, because of the percentage error of validated data can be considerable and the predicting power is good in the fitted model; the non-normality of residuals would not be a significant impact for the model.

Model equation for ARIMA (1,1,1) - GARCH (1,1)

Let Y_t be the original series. Then the 1st difference is of log series is: $\Delta \log(Y_t) = \log(Y_t) - \log(Y_{t-1})$

Thus, the mean equation: $\Delta log(Y_t) = 0.0007167 + 0.9150 \cdot \Delta log(Y_{t-1}) - 0.9409 \cdot \varepsilon_{t-1} + \varepsilon_t$

Variance equation: $\sigma_t^2 = 2.4146 \times 10^{-6} + 0.1206 \cdot \varepsilon_{t-1}^2 + 0.8612 \cdot \sigma_{t-1}^2$

Forecasting training data

Forecast: CLOSE_PRICF	
Actual: CLOSE_PRICE	
Forecast sample: 1/03/2000 4/02/2025	
Adjusted sample: 1/05/2000 4/02/2025	
Included observations: 6349	
Root Mean Squared Error	2.415763
Mean Absolute Error	1.391253
Mean Abs. Percent Error	0.805936
Theil Inequality Coefficient	0.005243
Bias Proportion	0.000531
Variance Proportion	0.000049
Covariance Proportion	0.999420

Figure 7: *Forecasting statistics (Train data)*

According to figure 7, Theil Inequality Coefficient is close to 0 (0.005) and bias proportion is also close to 0 (0.0010), it indicates the forecasting accuracy of the model is good. The Mean absolute percentage error (MAPE) is 0.8. These statistics show how accurate and strong the postulated model is.

Validation of the fitted model with the independent data set

Forecast: CLOSE_PRICF	
Actual: CLOSE_PRICE	
Forecast sample: 4/03/2025 6/13/2025	
Included observations: 50	
Root Mean Squared Error	26.33686
Mean Absolute Error	21.23720
Mean Abs. Percent Error	3.896595
Theil Inequality Coefficient	0.023088
Bias Proportion	0.124397
Variance Proportion	0.841029
Covariance Proportion	0.034574

Figure 8: *Forecasting statistics (Test data)*

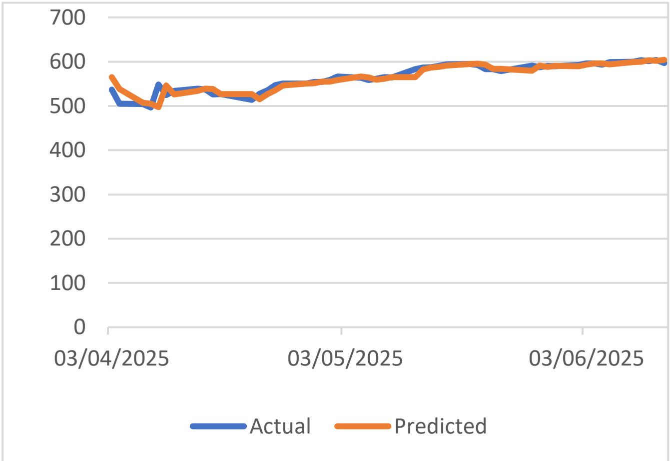

Figure 9: *Actual prices with predicted prices*

The fitted model ARIMA (1,1,1) - GARCH (1,1) was tested by an independent data set from 4/3/2025 to 13/06/2025 and compared with actual values. The error percentages between actual closing stock prices and predicted closing stock prices varies between -13.9% and 3.9%. with a mean absolute percentage error of 3.89. It implies that the model can be validated for the independent data set.

Table 6: *Predicted values for future days*

Date	Closing Stock Prices ($)
6/16/2025	603.9586874
6/17/2025	604.1258965
6/18/2025	604.3091358
6/19/2025	604.5073765
6/20/2025	604.719742
6/23/2025	604.9454058
6/24/2025	605.1835914
6/25/2025	605.4335694

6/26/2025	605.6946543
6/27/2025	605.9662018
6/30/2025	606.2476066

Table 6 indicates the predicted closing stock prices for upcoming few days (Except weekends and Holidays). These predictions were predicted by using the ARIMA (1,1,1) - GARCH (1,1) model and future decisions can be made based on the above forecasted stock price.

Conclusions

The fundamental aim of this study was to postulate a time series forecast of the daily long run movement of the SPDR S&P 500 ETF Trust (SPY) by applying the Box-Jenkins ARIMA approach, utilizing the data from January 2000 to June 2025. Most of the previous research has been centered solely around the index itself (S&P 500), but fewer studies have specifically targeted SPDR S&P 500 ETF Trust (SPY) despite the subtle but noticeable differences between the two. The best fitted model for the long run SPDR S&P 500 ETF Trust (SPY) is the ARIMA (1,1,1) + GARCH (1,1), with all coefficients being significant, lower AIC, SC and HQC values. Furthermore, the diagnostic tests confirmed the fact that the errors random and white noise.

The model was tested using an independent data set, in which the percentage errors varied from -13.9% to 3.9%, with the Mean Absolute Percentage Error being 3.89%, which manifests the accuracy and predictive power of the postulated model. The model is relatively simple and not overfitted, which helps people assess the long-term direction and behavior of the SPDR S&P 500 ETF Trust (SPY). The index seems to be on a growth trajectory, with years to come we predict the index could reach 606.25 $ by the end of June 2025.

When forecasting, however, past values should not be the sole basis upon which future predictions are made. It is also necessary to incorporate subjective aspects to improve the validity and accuracy of the forecast. As an example, on liberation day 2025 when the Trump administration announced major tariffs on some of the key trading partners with the U.S., the SPY plunged approximately 11% in just two days. This was mainly caused due to the investor sentiment and speculation in the market as a reaction to such one-off events. Hence it is necessary to address such changes by incorporating structural breaks into the time series model by considering dummy variables. By doing so, the model can better capture the underlying volatilities in the price movements of the SPDR S&P 500 ETF Trust (SPY), leading to more accurate and reliable forecasts.

The results of this research will be of immense benefit to policy makers, especially the FED in order to gauge financial market sentiment and systematic risks, both retail and institutional investors in order to position their portfolios and also corporate executives when deciding on investments that are worth undertaking with respect to the equity markets.

Acknowledgment

We are deeply thankful to the head of the Department of Mathematics for his unmatched support and encouragement throughout our research journey. We would also like to acknowledge the lecturers whose insights, knowledge and expertise helped us immensely in shaping and guiding our work to a successful completion.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Holden-Day.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987-1007. https://doi.org/10.2307/1912773

Kitchen, G. (2023). *S&P 500 (SPY) historical stock prices* [Data set]. Kaggle. https://www.kaggle.com/datasets/gkitchen/s-and-p-500-spy

Nickolas, S. (2024). SPY EFT: The SPDR S&P 500 ETF and what it holds. *Investopedia*. https://www.investopedia.com/articles/investing/122215/spy-spdr-sp-500-trust-etf.asp

State Street Global Advisors. (2025). *SPDR S&P 500 ETF Trust (SPY)*. https://www.ssga.com/us/en/intermediary/etfs/spdr-sp-500-etf-trust-spy