Title: Time Series Analysis of SPY ETF Prices: An In-Depth Examination of ARMA, ARCH, and ARMA-GARCH Models

Abstract: This report investigates the time series properties of daily adjusted closing prices for the SPY ETF using ARMA, GARCH, and ARMA-GARCH models. The analysis assesses the time-varying volatility and forecasts future observations. The study's findings will help stakeholders make informed decisions or predictions based on evidence provided by these models.

0. Github Repository

[Github Link](https://github.com/danielherrerahsph/SPYtimeseries)
https://github.com/danielherrerahsph/SPYtimeseries

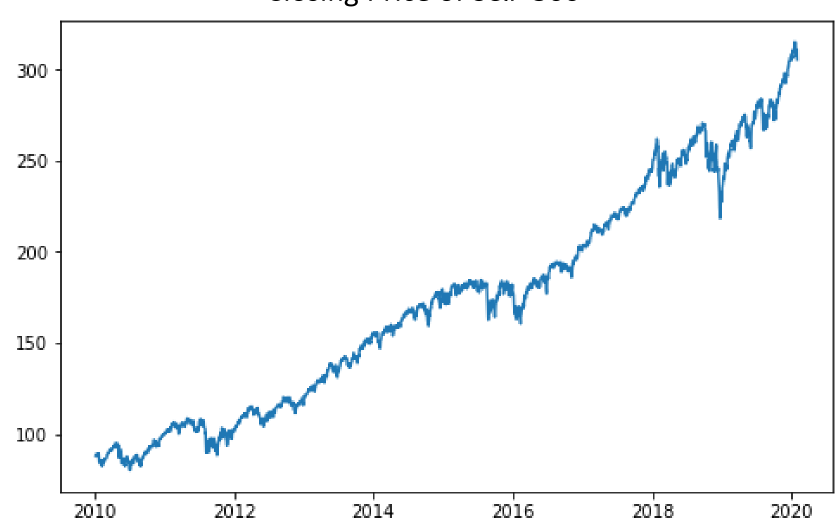
1. Introduction

The S&P 500 Index, a widely-followed equity benchmark, reflects the performance of 500 large-cap US companies spanning multiple sectors. The SPDR S&P 500 ETF Trust (SPY) is an exchange-traded fund (ETF) designed to track this index. In this study, we aim to model the time series behavior of daily adjusted closing prices or returns for the SPY ETF using ARMA, GARCH, and ARMA-GARCH models. The goal is to examine the time-varying volatility and predict future observations to aid stakeholders in making informed decisions.

2. Data Description

The dataset consists of daily adjusted closing prices for the SPY ETF obtained from Yahoo Finance using python's 'yfinance', a convenient wrapper around the Yahoo Finance API. The time range covers January 4, 2010, to February 1, 2020, resulting in a sample size of 2537 observations. The dataset is available in the file "spy_returns.csv" in the GitHub repository.

Closing Price of S&P 500



3. Time Series Models

The study employs three time series models:

a) Autoregressive Moving Average (ARMA) model: A combination of autoregressive (AR) and moving average (MA) processes, ARMA models the relationship between the current observation and previous observations, as well as the past errors.

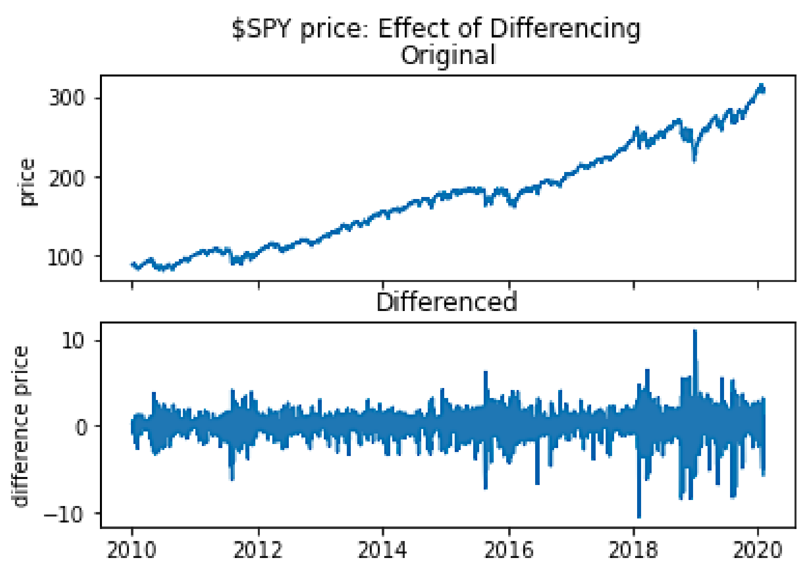
b) Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model: The GARCH model accounts for the time-varying volatility in the time series data by modeling the variance of error terms as a function of past errors.

c) ARMA-GARCH model: The ARMA-GARCH model combines ARMA and GARCH components, capturing both the temporal dependence and time-varying volatility in the time series data.

4. Preliminary Analysis

Before fitting the models, a preliminary analysis was conducted on the SPY ETF data:

a) Stationarity: A core assumption that had to be met for the time series modeling approaches was that the time series was stationary. In order to achieve this, differencing was used.

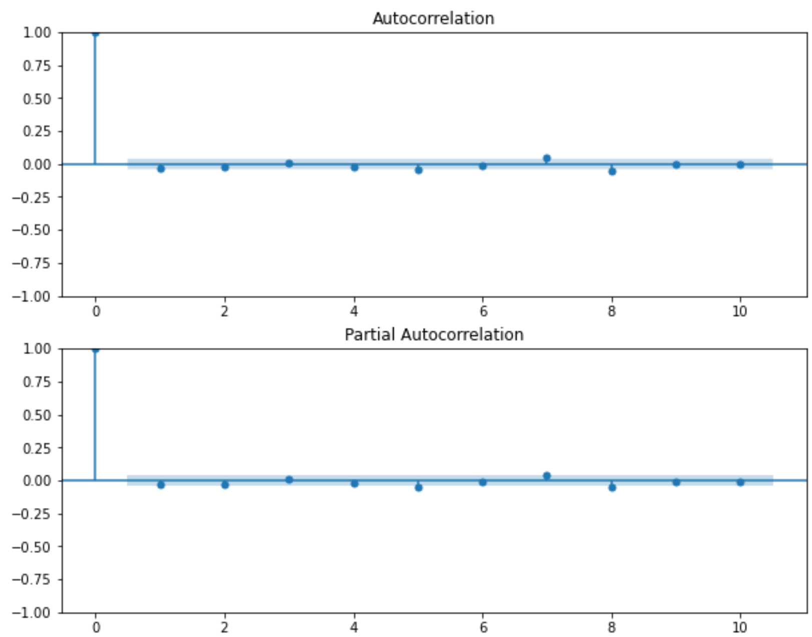


Additionally, the Augmented Dickey-Fuller (ADF) Test to test a null hypothesis that the series is not stationary versus the alternative hypothesis that the series is stationary. The test showed that the series was non-stationary, requiring differencing to achieve stationarity. After first-order differencing, the ADF test confirmed that the transformed series was stationary.

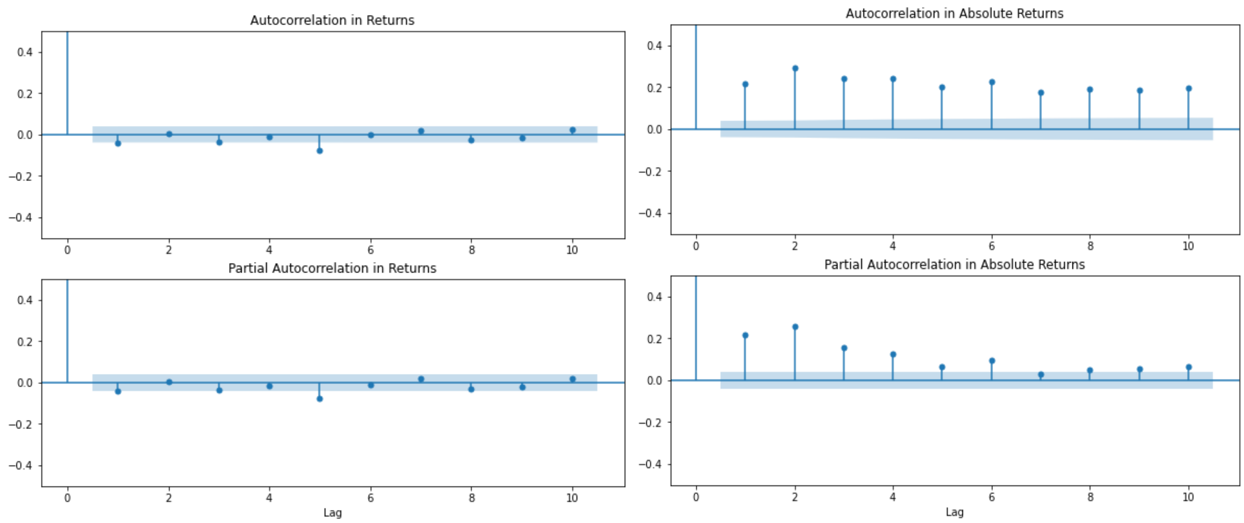
b) Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots: These plots were used to identify the order of the ARMA model using the following rules:

	AR(P)	MA(Q)	ARMA(P,Q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

The ACF and PACF plots were indicative of a random walk process, ARIMA(0,1,0) before differencing.



c) Log Returns: Subsequent analysis utilizing the volatility – GARCH and ARMA+GARCH – utilized the log returns rather than closing prices. Additionally, the absolute returns were used when testing assumptions and validity of these modeling processes due to the ability to handle outliers.



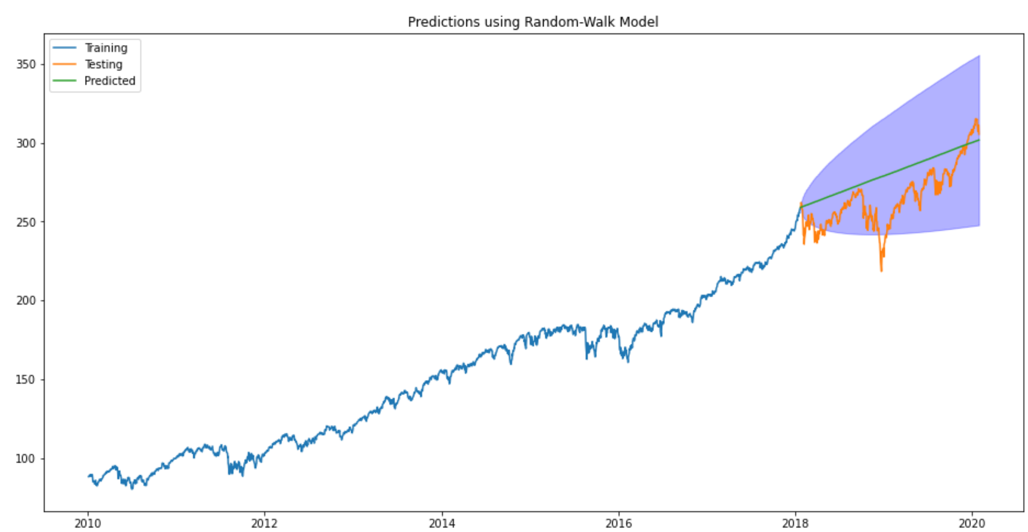
5. Estimated Models and Forecasts

The analysis results in the following models:

a) ARIMA(0,1,0), Random Walk

We fit the random walk process directly on the closing prices of the S&P 500.

$$Y_t - 0.0843 = W_t, where\ W_t \sim wn(1.5)$$



b) GARCH(1,1) model:

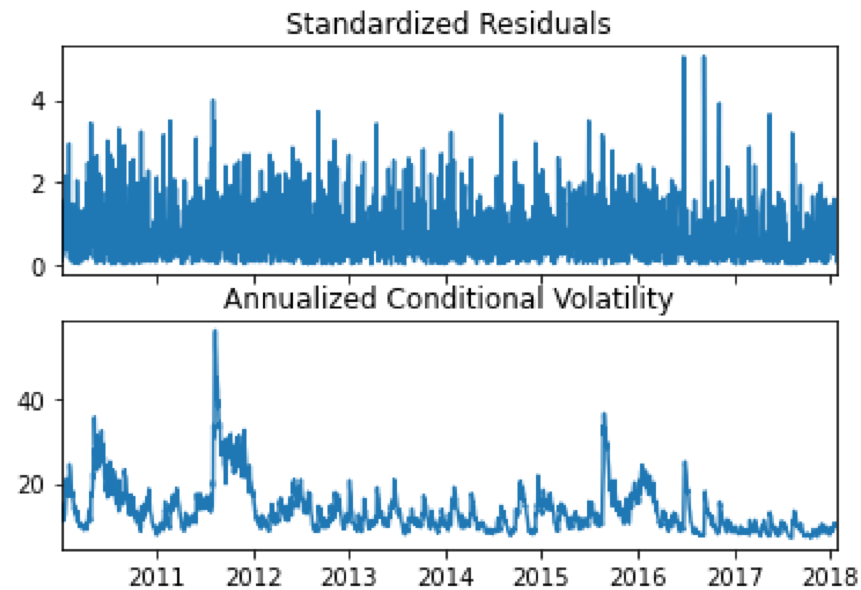
The ARCH model captures the volatility clustering in the log returns.

$$X_t = \sigma_t \epsilon_t$$

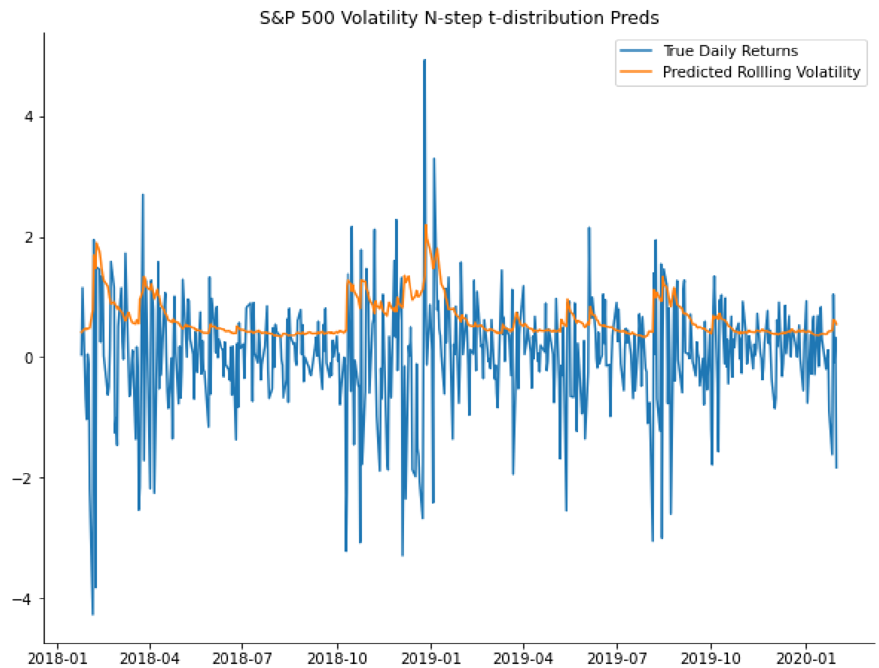
$$\sigma_t^2 = (1.7162 \times 10^{-6}) + 0.1X_{t-1}^2 + 0.88\sigma_{t-1}^2$$

where σ_t^2 is the conditional variance at time t, X_{t-1}^2 is the squared return at time t-1, and σ_{t-1}^2 is the conditional variance at time t-1. This assumes normally distributed error terms, however consideration and implementation with student's t-distribution will follow.

We notice that the GARCH(1,1) model has coefficients of 0.1 and 0.88 which sum to 0.98 and are very close to 1. This finding is expected and indicative of the potential modeling of this process with Integrated Generalized Autoregressive Conditional heteroskedasticity (IGARCH), a restricted version of the GARCH model, where the parameters sum up to one. This along with the ACF and PACF shown above for returns and absolute returns (see 4c) are indicative of long-range dependencies as observed by the noticeable non-zero autocorrelations for *k*-lags.



c) GARCH with rolling forecasts (one-step ahead):
A GARCH(1,1) model was fit to the time series using one step ahead predictors, as this might be more commonplace in the dynamic setting of trade strategy and trade execution. Thus, the fitted values changed as the training data expanded with each one day ahead inclusion of the observed values. This was done assuming both normal distribution in the errors and a t-distribution. The t-distribution rolling forecast is shown below.

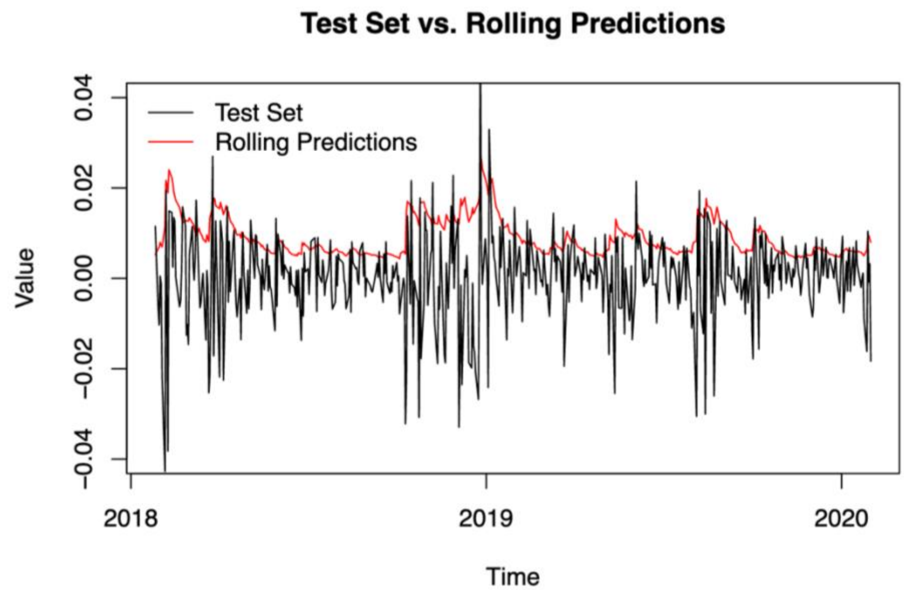


d) ARMA(1, 0) + GARCH(1, 1) model:

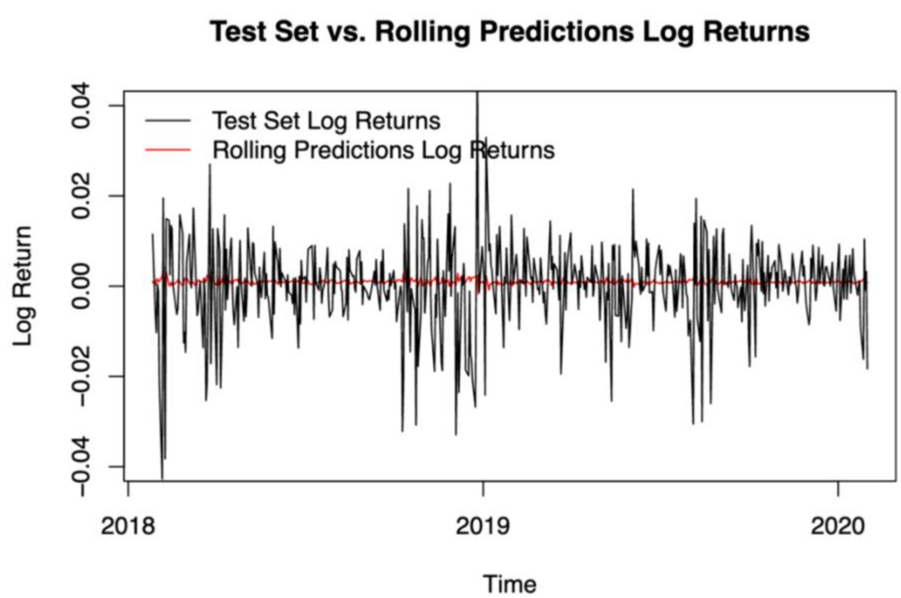
The ARMA-GARCH model captures both the temporal dependence and time-varying volatility in the log returns. The process below found an ARMA(1,0) model to have an AR(1) coefficient of -0.054. This term is small in magnitude, however interesting because it is statistically significant at the alpha 0.01 level (p-value = 0.008). Overall, this finding still implies that the GARCH effect is stronger than the autoregressive effect.

$$X_t = .001 - .054\, \mathrm{X}_{\mathrm{t-1}} + \mathrm{W}_{\mathrm{t}}$$
$$W_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = 2.587 \times 10^{-6} + .18\mathrm{W}_{\mathrm{t-1}}^2 + .81\sigma_{\mathrm{t-1}}^2$$

Similar to in c) above, the GARCH process was also fit using a rolling window in a subsequent analysis. The result of this process on forecasts can be shown below where we include the log returns of the test set and the rolling volatility predictions using the ARMA-GARCH process.



As a result of our findings that the coefficient in the AR(1) effect on the log returns is significant, we decide to plot the rolling predictions for the log returns and compare them to the test set log returns. We notice that when using a rolling window, as done with the volatility, the log returns are only trivially different than 0. One idea, however, is to see the direction that they are moving in and if that lines up with the direction of the test set; that is if the forecast for one day ahead is a positive return, is the test set also positive and likewise for negative. We see that the accuracy for the correct prediction of the direction of the movement of log returns is 0.56 or 56%.



6. Model Diagnostics and Validation

Model adequacy is assessed using different metrics for the aforementioned models. We will break down the diagnostics and validations procedures used for each modeling technique. The details of these diagnostics can be found in the respective notebooks for each modeling process.

a) ARIMA/ARMA: ADF, ACF, PACF, Ljung-Box test, normal QQ plot, residual diagnostics, MAPE, MSE

b) GARCH: ACF and PACF (returns and absolute returns), Ljung- Box test, McLeod-Li test, Jacque-Bera, ACF of Standardized Residuals, normal QQ plots

c) ARMA + GARCH: McLeod-Li tests

residual diagnostics, including the Ljung-Box test for serial correlation and the McLeod-Li test for ARCH effects. The ARMA-GARCH model emerges as the best fitting model, as it effectively captures both the short-term dependence and volatility clustering in the data.

7. Discussion

The ARMA-GARCH model demonstrates the ability to capture the dynamics of the SPY ETF returns effectively. However, there are some limitations and potential areas for improvement:

a) Model Complexity: While the ARMA-GARCH model is appropriate for this analysis, more sophisticated models could potentially provide better results. Additionally, incorporating external factors, such as macroeconomic variables or market sentiment, may improve the model's explanatory power.

b) Data Preprocessing: Analyzing the time series at different frequencies (e.g., weekly or monthly) might provide different insights into the data.

c) Model Selection: Rigorous model selection techniques, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), could help identify the optimal combination of ARMA and GARCH components for the model, something that was not explored fully due to time.

ARMA-GARCH model captures both the temporal dependence and time-varying volatility in the data and provides insights into potential future price movements and volatility. The report's findings can aid stakeholders in making informed decisions or predictions based on evidence provided by these models. These models can be combined with trading strategies and backtesting to implement different strategies. Next steps are to combine this strategy with a threshold for volatility at which to execute certain trades. Back testing will be used to optimize a strategy. A wrapper around the Robinhood API will be used to execute trades.

If directly implementing trading strategies directly using the volatility are not successful, there is still use for the volatility modeling. For example, the volatilities can be utilized as features for more advanced models such as deep learning or other machine learning models.