Modeling S&P500 Volatility Clustering using ARIMA-GARCH

Date: 24-04-2025

1 Introduction

Financial market volatility, representing the magnitude of asset price fluctuations, is a fundamental concept influencing risk management, derivative pricing, and investment strategies. A well-established characteristic of financial returns is volatility clustering, where large price changes tend to be followed by further large changes, and small changes are followed by small changes. This phenomenon, termed time-varying conditional heteroscedasticity, violates the constant variance assumption of basic linear time series models, necessitating more sophisticated approaches. This study investigates volatility dynamics within the S&P 500 index (gspc), a primary benchmark for the U.S. equity market. The central hypothesis is that S&P 500 daily returns exhibit significant volatility clustering, which can be effectively captured and modeled by a GARCH(1,1) process following an ARIMA(1,1) filter for the conditional mean. This hypothesis stems directly from the observed empirical regularities in financial markets and the development of the Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized ARCH (GARCH) models specifically designed to address such patterns. The analysis leverages a substantial dataset of daily S&P 500 observations spanning from 2004 to 2025, providing ample data to identify and model these persistent volatility dynamics. To provide context and further validation, secondary analyses examine the relationship between the estimated S&P 500 volatility and both the CBOE Volatility Index (vix, a measure of market-implied volatility) and S&P 500 trading volume, as well as modeling the dynamics of the VIX itself. The primary dataset for this analysis consists of daily financial time series obtained from Yahoo Finance using the yfinance Python library. The specific series utilized are the S&P 500 index (gspc) Adjusted Close prices and trading Volume, and the CBOE Volatility Index (vix) Closing prices. The data spans the period from January 1, 2004, to April 23, 2025. Prior to analysis, the time series were aligned to a common daily index using pandas, ensuring that all calculations and comparisons were performed on days where data was available for all relevant series. This alignment process resulted in a final dataset containing 5360 concurrent daily observations. S&P 500 Adjusted Close prices were transformed into daily logarithmic returns for the core volatility modeling, while VIX levels and S&P 500 Volume (and its logarithm) were used for comparative analyses.

2 Methodology

This study employed a multi-step methodology rooted in standard financial econometrics practice for analyzing time series volatility. First, S&P 500 Adjusted Close prices were transformed into daily logarithmic returns ($\log(P_t/P_{t-1}) \times 100$) to achieve stationarity and work with percentage changes. Stationarity of the transformed S&P 500 series and the raw VIX level series was formally assessed using the Augmented Dickey-Fuller (ADF) test, which tests the null hypothesis of a unit root (non-stationarity). The results of the ADF test guided the choice of the integration order (d) for subsequent ARIMA modeling. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots served as crucial diagnostic tools. They were examined for the S&P 500 log returns and VIX levels to identify patterns of linear dependence, aiding in the selection of appropriate Autoregressive (p) and Moving Average (q) orders for the ARIMA models. More importantly, ACF/PACF plots of the squared log returns and squared model residuals were analyzed to visually detect the presence and persistence of autocorrelation, which is the signature of ARCH effects (volatility clustering). For the VIX series, the Ljung-Box test was additionally applied to the squared series to statistically corroborate the visual evidence of ARCH effects. An ARIMA model was employed to capture the linear dependencies in the conditional mean of each relevant series (gspc log returns, vix level). The purpose of this step was to filter out any significant linear autocorrelation, ensuring that the residuals passed to the volatility model represent primarily the unpredictable component of the returns or levels. Based on stationarity tests and ACF/PACF diagnostics, an ARIMA(1,0,1) specification (equivalent to ARMA(1,1) for the stationary log returns) was selected for the S&P 500, and an ARIMA(1,0,1) was chosen for the VIX level. This choice reflects common practice for financial returns (often weak ARMA structures) and persistent indices like VIX (strong ARMA structure). While formal model selection criteria like AIC/BIC could be used for order selection, this study utilized the standard (1,1) specifications. The core of the volatility analysis involved fitting a GARCH(1,1) model to the residuals obtained from the ARIMA estimation step for both the S&P 500 and VIX series. The GARCH model directly addresses the time-varying conditional variance by specifying it as a function of past squared residuals (the ARCH term, reflecting reaction to past shocks) and past conditional variances (the GARCH term, reflecting volatility persistence). The GARCH(1,1) specification, with one ARCH and one GARCH lag, was chosen as it is a parsimonious yet often highly effective choice for capturing the stylized facts of financial volatility and justified by the observed ARCH effects. Validation of the fitted GARCH models relied principally on analyzing the standardized residuals ($\text{residual}_t/\sqrt{\text{conditional\_variance}_t}$). Specifically, ACF/PACF plots of the squared standardized residuals were examined; if the GARCH model successfully captures the conditional heteroscedasticity, these plots should show no significant remaining autocorrelation, resembling white noise. ACF/PACF plots of the non-squared standardized residuals were also checked to confirm the adequacy of the initial ARIMA mean specification. Finally, Pearson correlation coefficients (R) were calculated to quantify the linear associations between the estimated S&P 500 GARCH volatility (annualized for VIX comparison, daily for volume comparison) and the VIX index level, and between the daily S&P 500 GARCH volatility and the logarithm of S&P 500 trading volume. Log volume was used to mitigate the influence of extreme volume outliers. Out-of-sample forecasts were generated from the fitted models using the respective `forecast` methods to illustrate potential applications. Visualization using `matplotlib` and `seaborn` (including time series, ACF/PACF, residual, forecast, and hexbin plots) was integral throughout the process. All analyses were performed using Pycharm and relevant scientific libraries (`pandas`, `numpy`, `yfinance`, `statsmodels`, `arch`, `scipy`, `matplotlib`, `seaborn`).

3 Exploratory Data Analysis (EDA)

Initial visual inspection of the S&P 500 adjusted-close price (Figure 1) revealed a clear upward trend over the 2004–2025 period, confirming non-stationarity in the price level and the need for transformation before modelling. Significant market downturns that correspond to major economic events are also clearly visible.

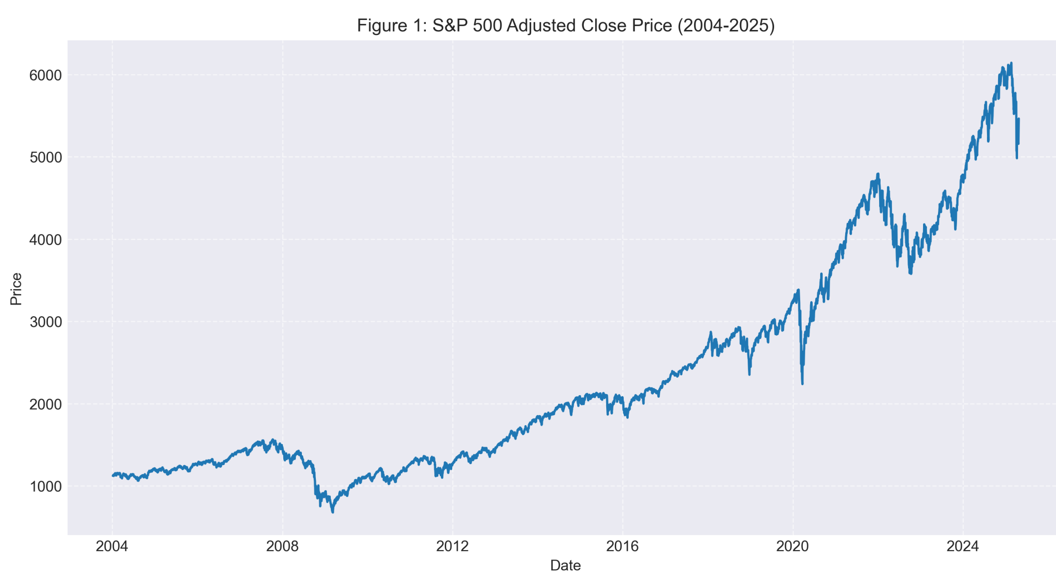

Figure 1: S&P 500 Adjusted-Close Price (2004–2025)

The daily log-returns of the S&P 500, plotted in Figure 2, appear stationary, fluctuating around zero without a discernible trend. This visual assessment is supported by the ADF test result (ADF statistic $= -18.6441$, p-value $= 0.0000$). Crucially, Figure 2 shows clear evidence of volatility clustering: periods of large positive and negative returns are grouped together (e.g. 2008–2009, 2020, early 2025), interspersed with periods of relative tranquillity. The ACF and PACF plots of the returns themselves (Appendix Figure 13) indicated minimal significant linear autocorrelation, suggesting an ARMA(1,1) is sufficient for the mean equation.

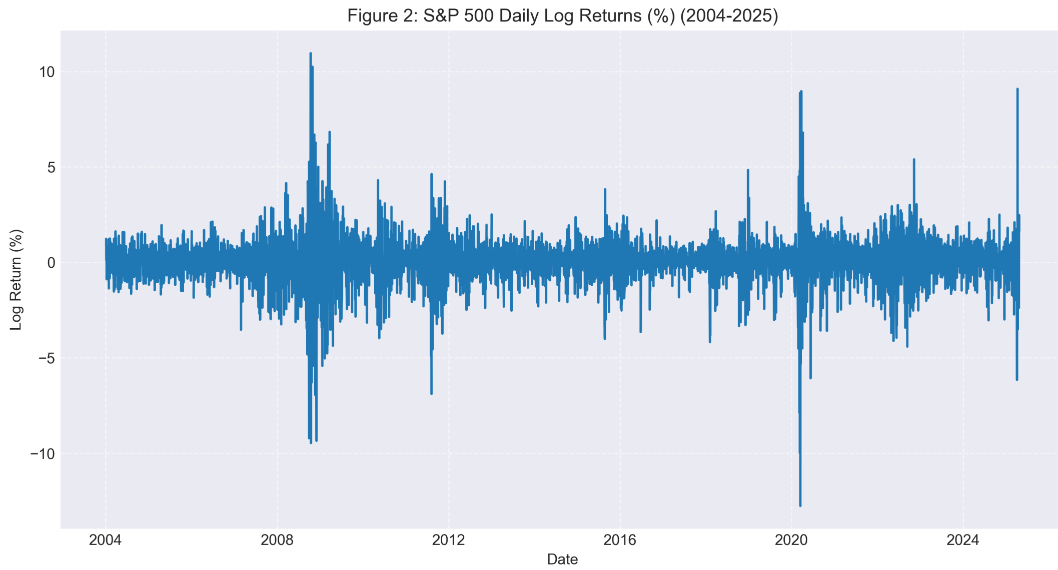

Figure 2: S&P 500 Daily Log-Returns (%) (2004–2025)

The key EDA finding supporting the primary hypothesis comes from the autocorrelation analysis of the *squared* log-returns, shown in Figure 3. The ACF plot exhibits significant positive correlations that decay very slowly, while the PACF shows significant values for the initial lags. This pattern is the definitive signature of ARCH/GARCH effects, confirming the presence of time-varying, persistent volatility clustering in S&P 500 returns and strongly motivating the use of a GARCH model. Similar EDA performed on the VIX index (Appendix Figure 11) confirmed its stationarity (ADF $p = 0.0000$) and the presence of ARCH effects (Ljung–Box test on squared series $p = 0.0000$, Appendix Figure 17), justifying the application of similar modelling techniques to the VIX as well.

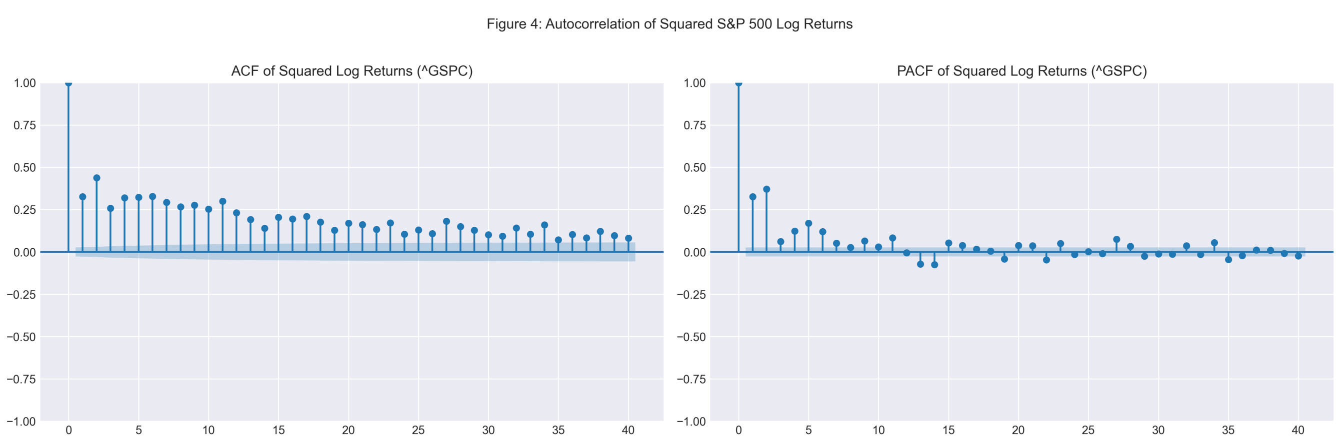

Figure 3: Autocorrelation of Squared S&P 500 Log-Returns

4 Model Specification and Estimation

Following the EDA, an ARIMA(1,0,1) model (effectively an ARMA(1,1) because $d = 0$ per the ADF test) was fitted to the stationary S&P 500 log-returns to capture conditional-mean dynamics. The resulting model summary showed statistically insignificant coefficients for the AR(1) ($p = 0.315$) and MA(1) ($p = 0.073$) terms at the 5 % level-consistent with the weak linear dependence observed in the log-return ACF/PACF plots. Despite the lack of significance, this ARMA structure serves as a preliminary filter before modelling the variance. Ljung-Box tests on the ARIMA residuals confirmed no remaining serial correlation ($\text{Prob}(Q) = 1.00$) but did reveal significant heteroscedasticity ($\text{Prob}(H) = 0.00$), reinforcing the need for a GARCH component.

The residuals from this ARIMA(1,0,1) fit were then modelled using a GARCH(1,1) specification. All estimated GARCH parameters were highly significant ($p < 0.001$): the constant $\omega = 0.0261$, the ARCH term $\alpha_1 = 0.1270$, and the GARCH term $\beta_1 = 0.8517$. The sum $\alpha_1 + \beta_1 = 0.9787$ quantifies the persistence of volatility shocks. Being close to-but less than-one implies a highly persistent yet stationary variance process, in line with the typical behaviour of equity-index volatility and directly supporting the primary hypothesis.

For the comparative VIX analysis, an ARIMA(1,0,1) model was fitted to the stationary VIX level series. This model exhibited highly significant AR(1) (coefficient $= 0.9832$, $p < 0.001$) and MA(1) (coefficient $= -0.1580$, $p < 0.001$) terms, capturing the strong persistence identified in the VIX EDA. A GARCH(1,1) model fitted to these ARIMA residuals also yielded significant parameters ($\omega = 0.1712$, $\alpha_1 = 0.3369$, $\beta_1 = 0.6631$, all $p < 0.001$). Notably, the persistence for the VIX GARCH model was $\alpha_1 + \beta_1 = 1.000$, indicating an integrated GARCH (IGARCH) process and implying virtually non-decaying volatility shocks for the VIX index itself.

5 Results

The primary result of this analysis is the successful fitting and validation (see Section 6) of the ARIMA(1,1)–GARCH(1,1) model to the S&P 500 log returns. This outcome supports the central hypothesis that this framework effectively captures the observed volatility clustering in the market index. The statistical significance of the GARCH parameters and the high persistence estimate (0.9787) provide quantitative evidence for the time varying and persistent nature of S&P 500 volatility.

Secondary analyses of other market variables yielded significant findings. A comparison between the annualised S&P 500 GARCH volatility and the VIX index level revealed a very strong positive correlation ($R = 0.8944$, $p = 0.0000$), as shown in Figure 4. This confirms that volatility estimated from historical price movements closely mirrors the market's forward looking expectation of volatility derived from option prices.

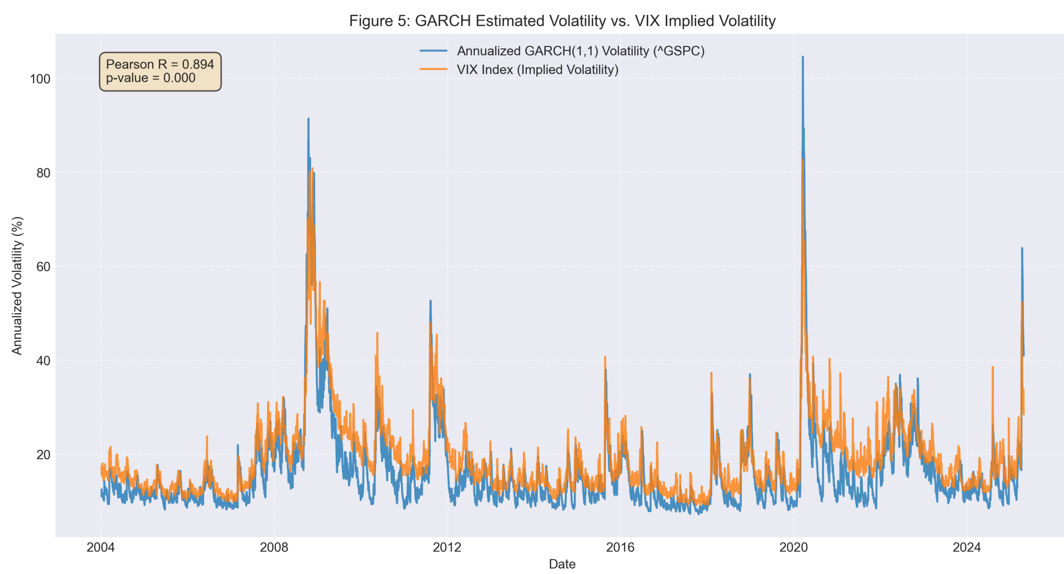

Figure 4: GARCH Estimated Volatility vs. VIX Implied Volatility

Examining the link between volatility and market activity, a moderate positive correlation ($R = 0.4579$, $p = 0.0000$) was found between daily S&P 500 GARCH volatility and the logarithm of S&P 500 trading volume. The hexbin plot in Figure 5 shows that higher volatility days generally coincide with higher log volume, although the relationship is noisy and most observations cluster in a low volatility, high log volume region.

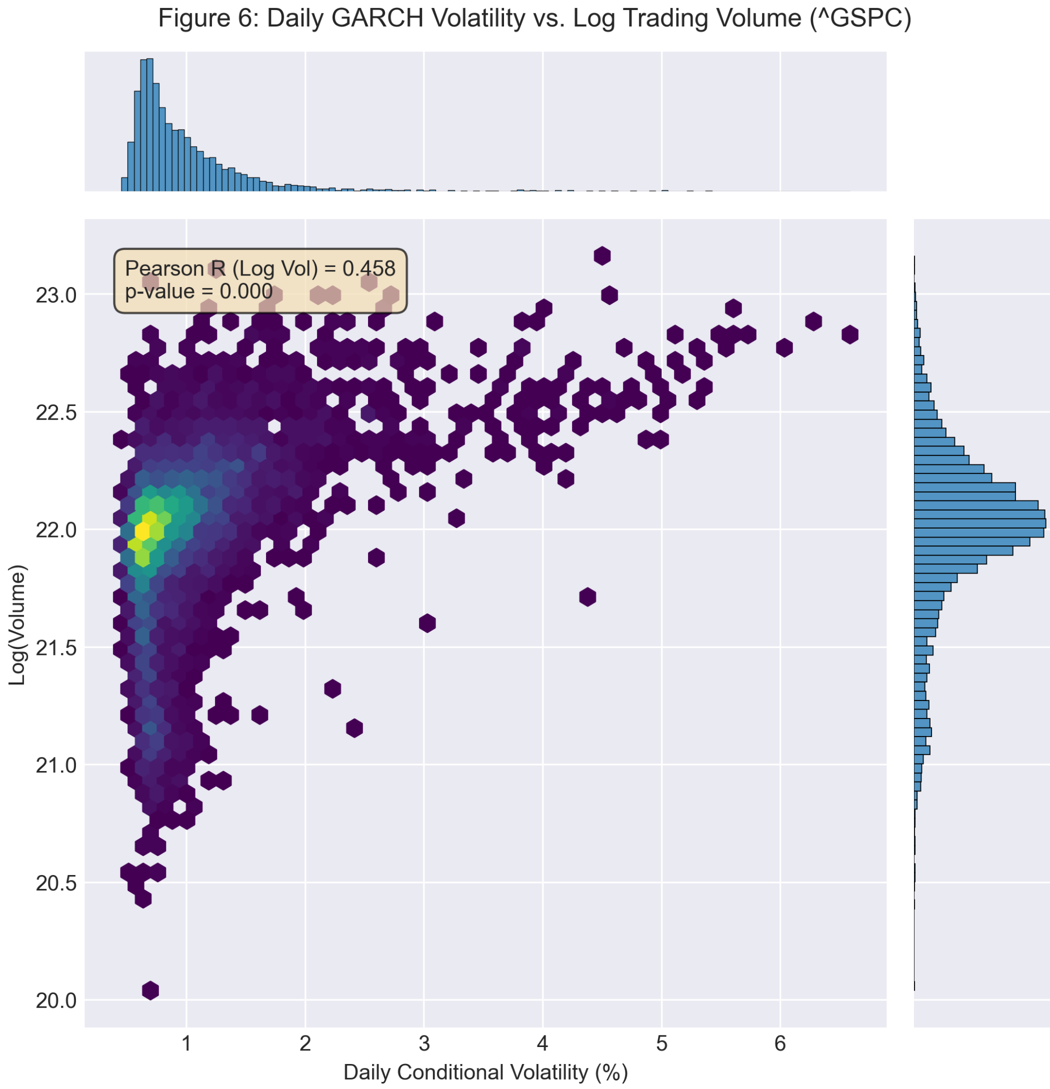

Figure 5: Daily GARCH Volatility vs. Log Trading Volume (S&P 500)

Finally, the models were used to generate out of sample forecasts. The forecast for S&P 500 conditional volatility is shown in Figure 6. It displays the characteristic mean reverting behaviour expected from a stationary GARCH process, with the forecast gradually decaying from the last observed volatility level toward the model's implied long run average. Forecasts for the VIX level and its conditional volatility were also produced (Figures 7 and 8), demonstrating the potential predictive applications of the fitted models.

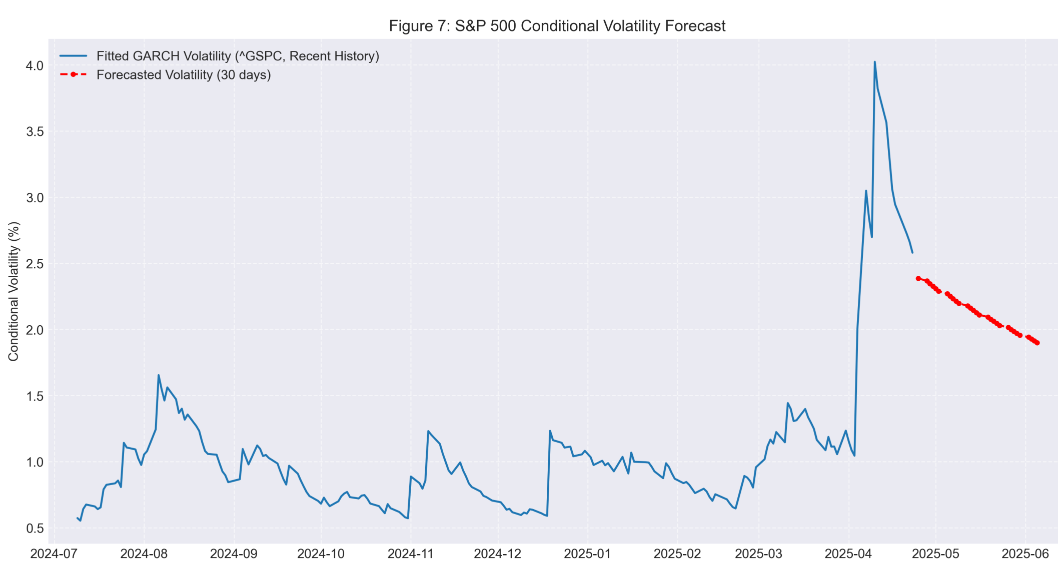

Figure 6: S&P 500 Conditional Volatility Forecast

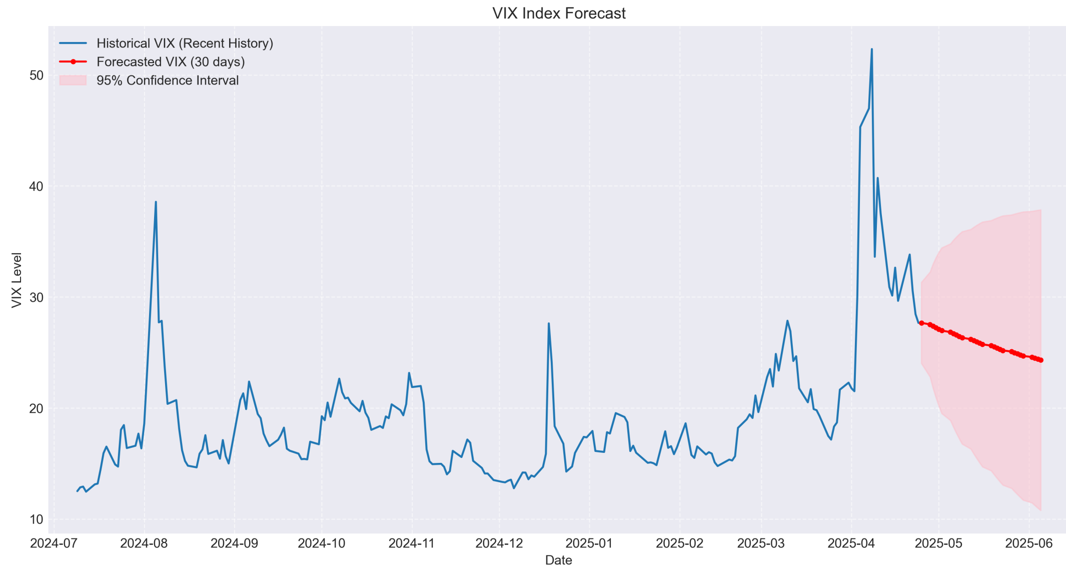

Figure 7: VIX Index Level Forecast

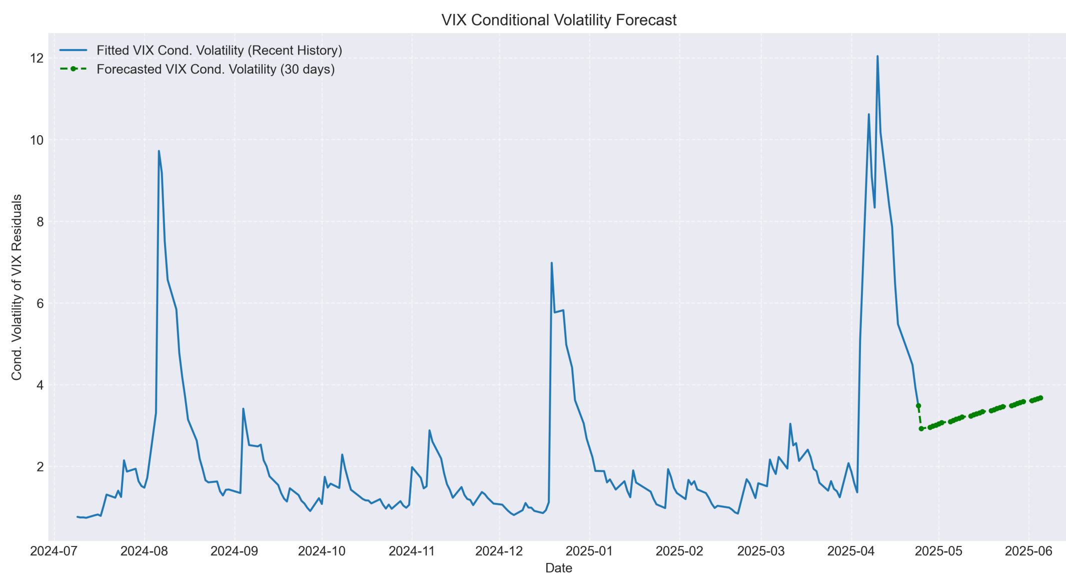

Figure 8: VIX Conditional Volatility Forecast

6 Model Validation

Robust validation is essential to confirm that the chosen ARIMA(1,1) − GARCH(1,1) model adequately represents the data generating process, especially the complex volatility dynamics. Assessment relied primarily on diagnostic checks of the model's standardised residuals. The crucial step for validating the GARCH component involves examining the autocorrelation structure of the *squared* standardised residuals. Figure 9 shows the ACF and PACF plots for these residuals from the fitted S&P 500 ARIMA(1,1) − GARCH(1,1) model.

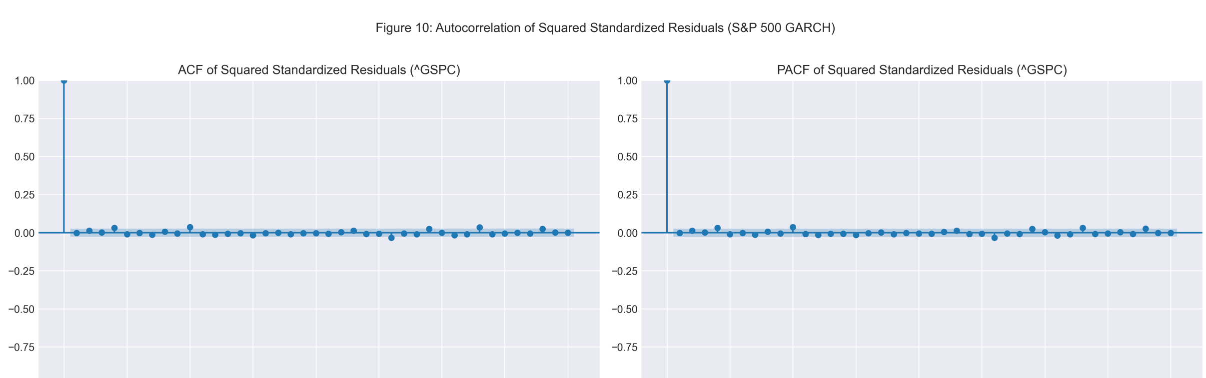

Figure 9: Autocorrelation of Squared Standardised Residuals (S&P 500 GARCH)

Figure 9 shows no evidence of significant autocorrelation in the squared standardised residuals.

All correlation coefficients at each lag fall within the 95 % confidence limits, matching the pattern expected from a white-noise process. This indicates that the GARCH(1,1) model has successfully captured the conditional heteroscedasticity (volatility clustering) that was visible in the squared raw returns (Figure 3). Analysis of the unsquared standardised residuals (Appendix Figure 15) also shows no significant autocorrelation, confirming that the ARIMA(1,1) component handled the conditional-mean specification adequately. Similar diagnostic checks for the VIX ARIMA–GARCH model likewise indicate a good fit (Appendix Figures 18 and 19).

Although these in-sample diagnostics are positive, a complete validation would ideally include an assessment of forecasting accuracy on a dedicated out of sample test set using suitable error metrics (for example, RMSE and MAE for level forecasts, or QLIKE and MSE for volatility forecasts). Such tests were not carried out in this study.

7 Discussion and Conclusions

This study confirmed the presence of volatility clustering in S&P 500 daily returns and showed that an ARIMA(1,1) GARCH(1,1) model captures this dynamic, validating the primary hypothesis. The high persistence estimate from the GARCH parameters ($\alpha_1 + \beta_1 \approx 0.98$) indicates that volatility shocks are long-lasting and must be considered in risk modeling.

The strong positive correlation between GARCH estimated volatility and the VIX index ($R \approx 0.89$, Figure 4) provides empirical support that historical volatility patterns mirror market expectations embedded in option prices. A moderate positive correlation between volatility and log trading volume ($R \approx 0.46$, Figure 5) further links market uncertainty with activity, though the hexbin plot suggests this relation is driven mainly by low-volatility, high-volume regimes contrasted with higher volatility periods that show more dispersed volume.

The fitted GARCH model yields dynamic estimates of S&P 500 volatility (Figure 6) that are directly useful for risk management tasks such as Value at Risk and dynamic asset allocation. The same approach can be applied to other financial time series with similar volatility behaviour.

Challenges included maintaining data alignment and handling intermittent `NaN` values. Dense data required plots such as hexbins for clarity (Figure 5). The near-IGARCH result for the VIX cautions that shock persistence may be effectively infinite. Limitations of this study include using a standard GARCH(1,1) rather than models that allow for asymmetry.

In summary, S&P 500 volatility is both clustered and persistent. An ARIMA GARCH framework captures these features well, as confirmed by residual diagnostics (Figure 9). Estimated volatility aligns closely with market-implied measures such as the VIX and shows a positive link with trading volume. The ARIMA GARCH approach therefore remains a valuable tool for understanding and quantifying risk in financial markets.

A Appendix Figures

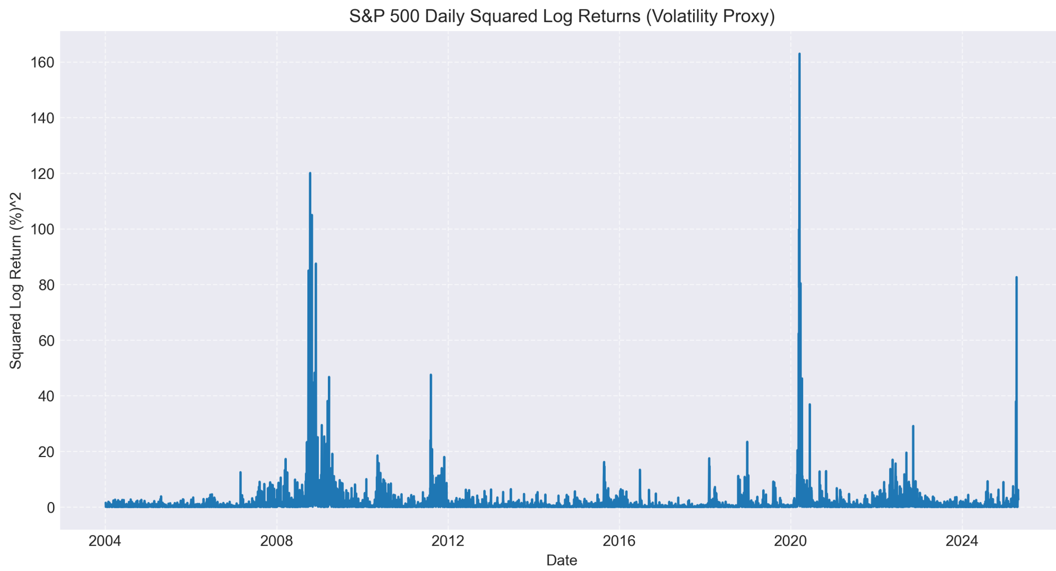

Figure 10: S&P 500 Daily Squared Log Returns (Volatility Proxy)

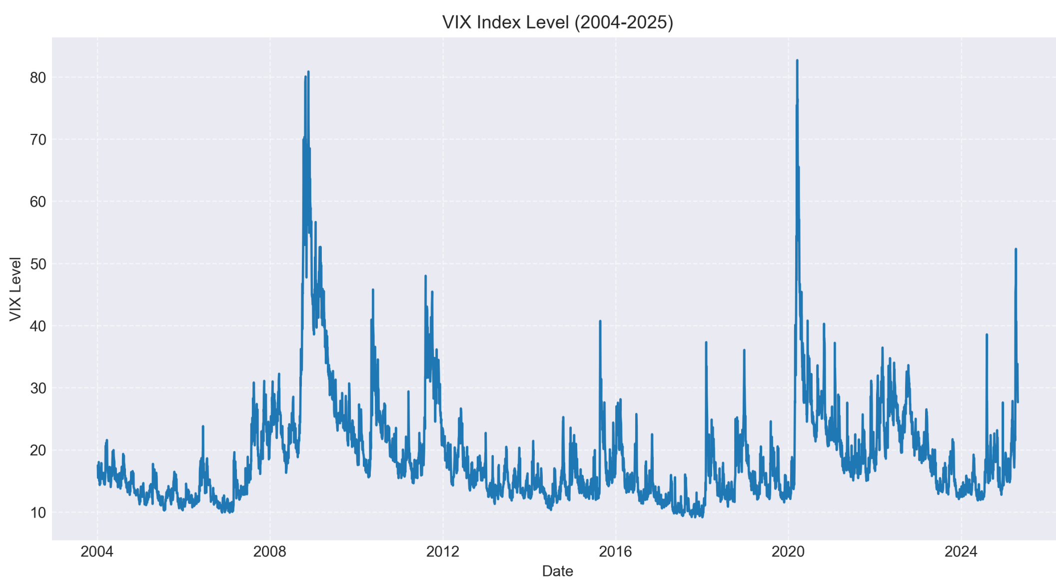

Figure 11: VIX Index Level (2004–2025)

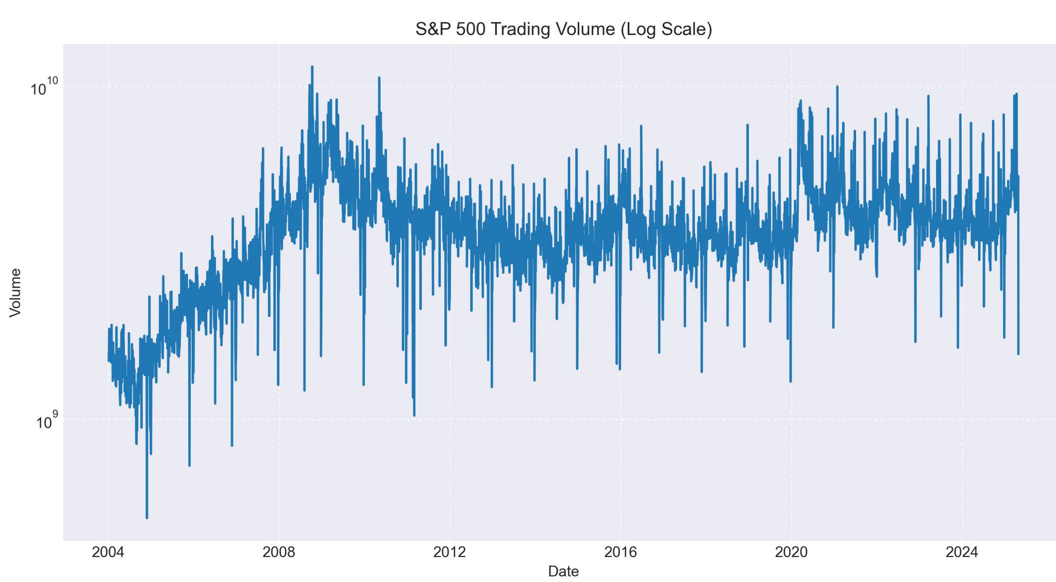

Figure 12: S&P 500 Trading Volume (Log Scale)

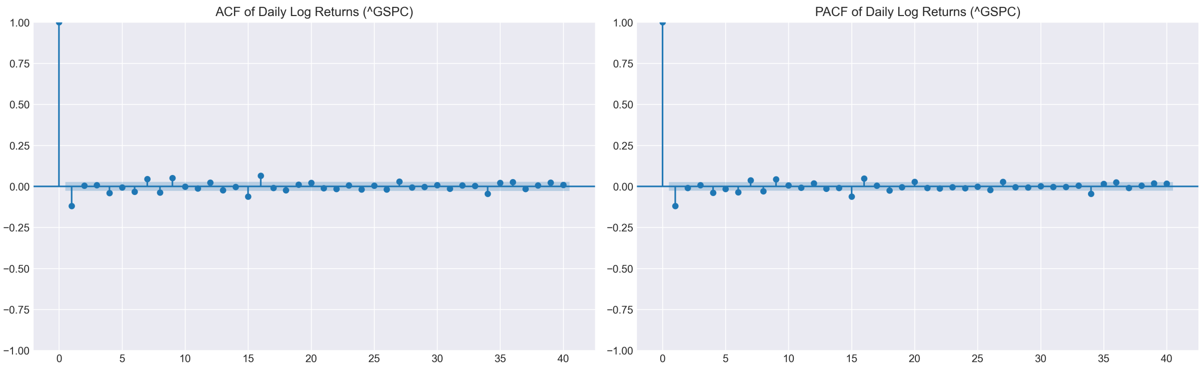

Figure 13: ACF/PACF of Daily Log Returns (S&P 500)

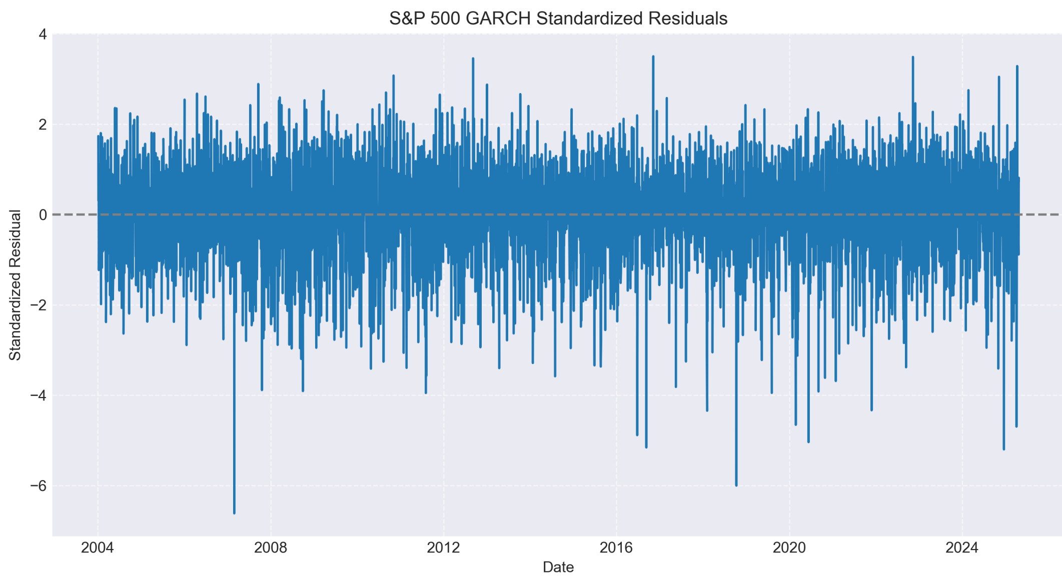

Figure 14: S&P 500 GARCH Standardised Residuals

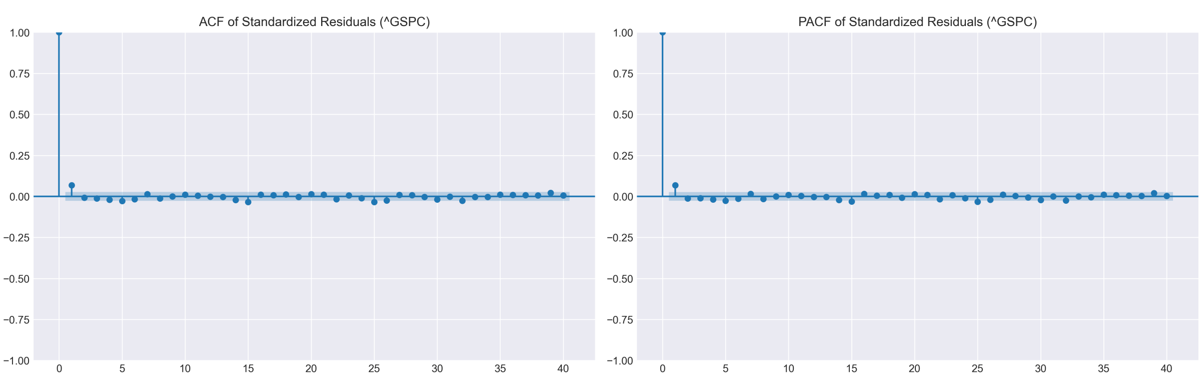

Figure 15: ACF/PACF of Standardised Residuals (S&P 500 GARCH)

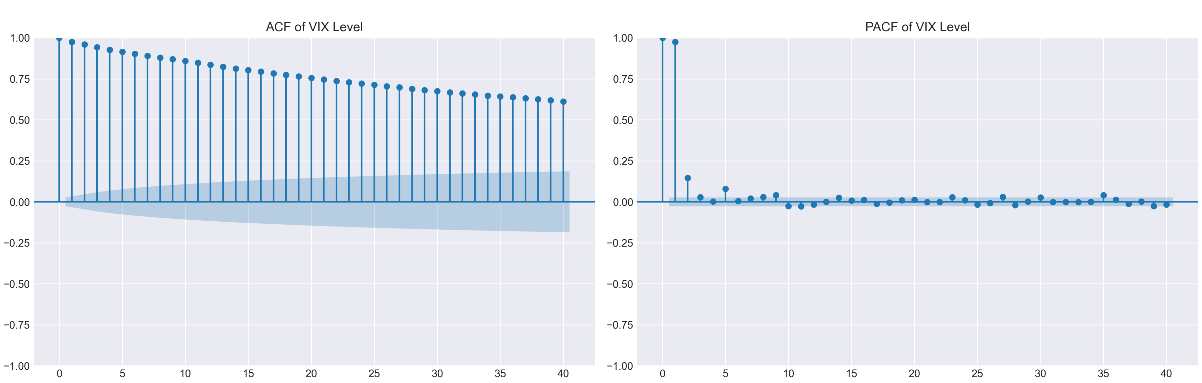

Figure 16: ACF/PACF of VIX Level

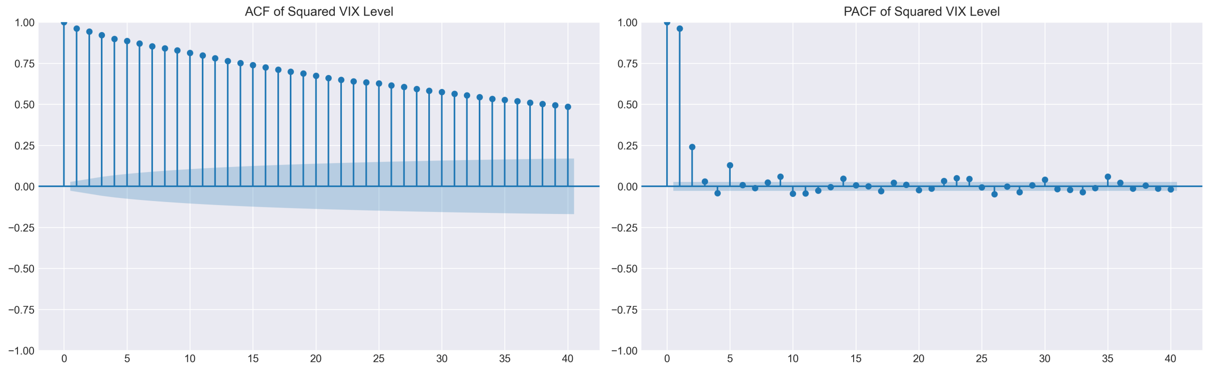

Figure 17: ACF/PACF of Squared VIX Level

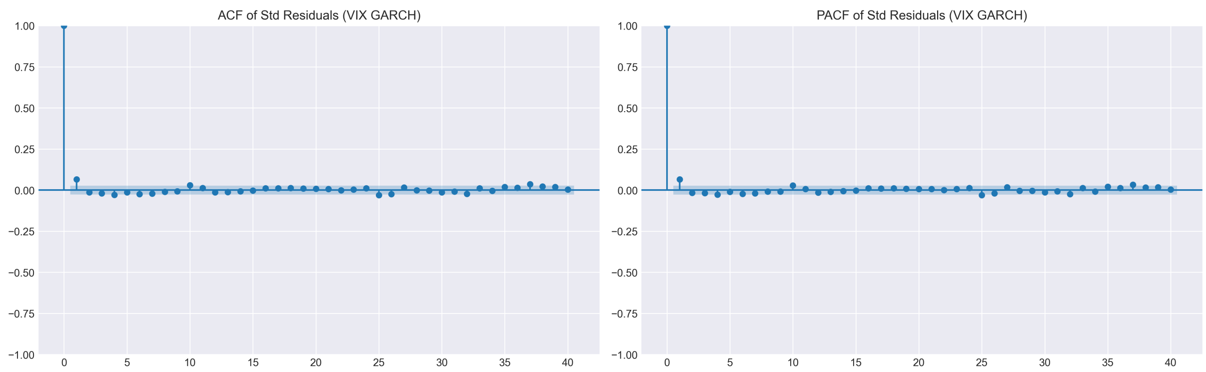

Figure 18: ACF/PACF of Standardised Residuals (VIX GARCH)

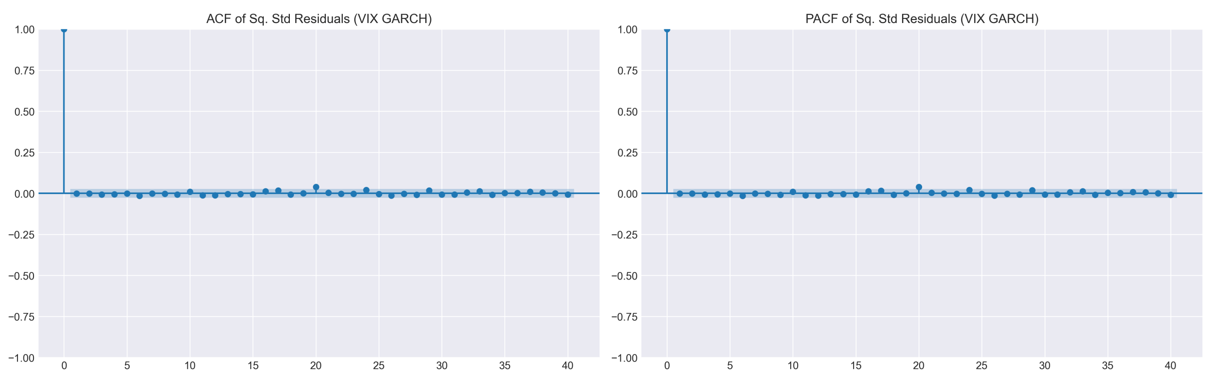

Figure 19: ACF/PACF of Squared Standardised Residuals (VIX GARCH)