GARCH Volatility Analysis - S&P 500

Ian Moore

2025-08-06

GARCH Volatility Modeling

Analysis of S&P 500 Volatility Dynamics

This notebook demonstrates advanced volatility modeling techniques using GARCH models applied to S&P 500 data. We progress from basic univariate GARCH to a multivariate model and rolling forecast. Lastly, we'll compile for risk management applications.

Key Analyses:

1. Univariate GARCH modeling and diagnostics
2. Dynamic Conditional Correlation (DCC-GARCH) for multi-asset portfolios
3. Rolling Window Forecasts
4. Risk management applications (VaR, portfolio optimization)

```
Daily S&P 500 returns were loaded for the period from 2020-01-01 to 2024-12-31.
Additional tickers were loaded for the same period: TLT, GLD, and VXX. These will be
used in the multivariate analysis.
```

1. S&P Summary Statistics and Exploratory Data Analysis

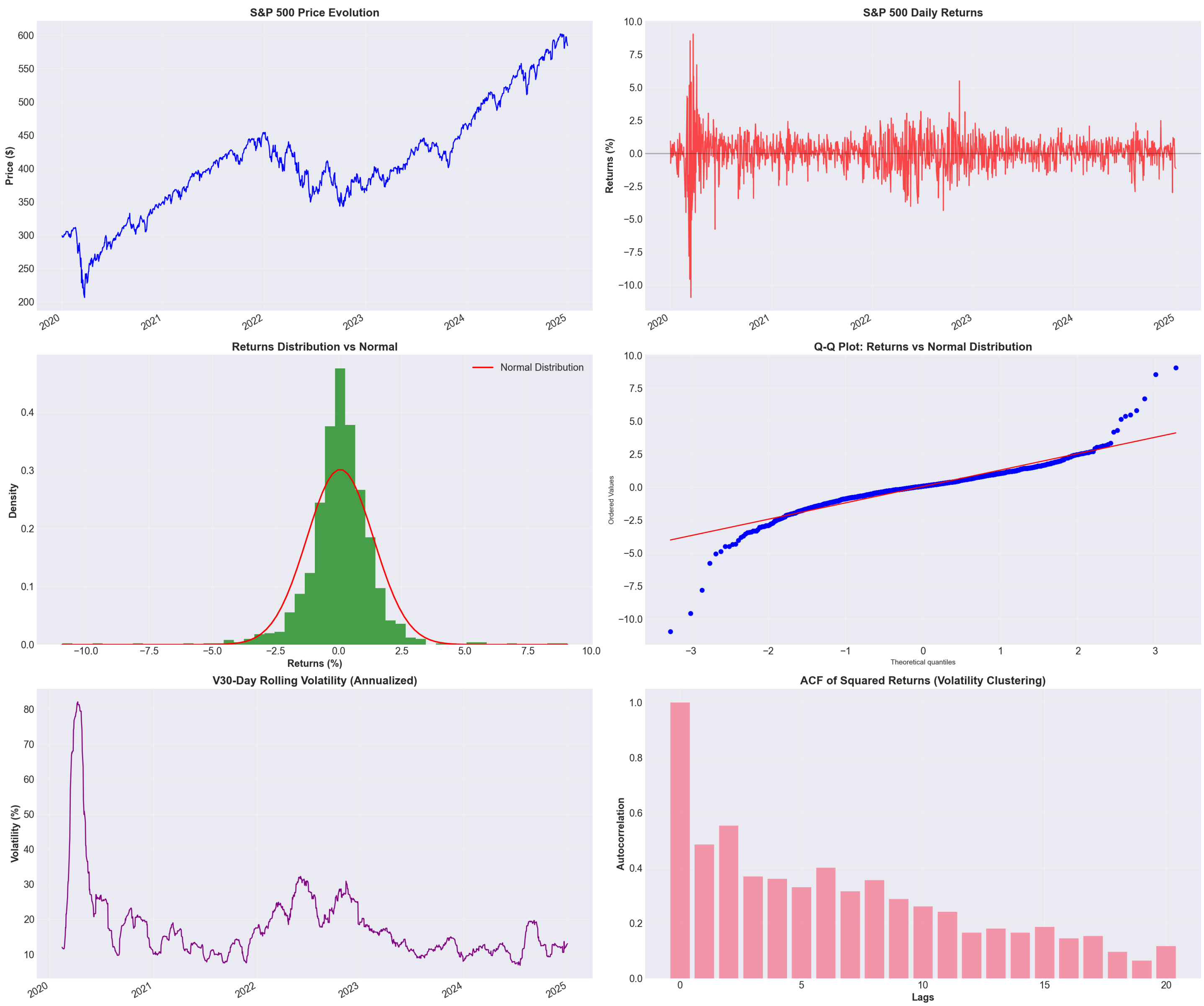
1.1 S&P 500 Summary Statistics and Analysis

```
S&P 500 summary statistics results are the following:

Mean: 0.0628%
Std Dev: 1.3229%
Skewness: -0.5443
Kurtosis: 11.5024
Min: -10.9424%
Max: 9.0603%
```

S&P 500 analysis of summary statistics:

The S&P 500 daily returns resemble the characteristics of typical financial time series data, with an average daily return of 0.0628%, or approximately 16% annualized. The standard deviation of 1.32% represents moderate daily volatility, while the negative skewness of -0.54 reveals a slight tendency toward larger negative returns than positive returns. The excess kurtosis of 11.50 demonstrates significant fat tails as the baseline kurtosis is 3.0, resulting in an excess of 8.50. This high kurtosis, combined with negative skewness (-0.54), indicates the S&P 500 has frequent small gains, occasional large losses, and reflects extreme events occurring more frequently than normal distribution theory would predict. This is evidenced by the wide range between the observed minimum and maximum from -10.94% to +9.06%. Finally, this confirms the presence of volatility clustering and suggests that traditional risk models assuming normal distributions may underestimate tail risks for this period of the portfolio.

1.2 Exploratory Data Analysis and Visualizations



S&P 500 price evolution shows an increase of nearly 200% from $300 to around $600. While demonstrating strong long-term growth, the data reveals two distinct periods of significant volatility clustering: early-to-mid

2020 during the COVID market crash, and throughout most of 2022 amid aggressive interest rate increases. This clustering behavior is clearly visible in the daily returns plot, where periods of high volatility (large price swings) are followed by continued high volatility, and calm periods persist for extended timeframes. The returns distribution and Q-Q plot confirm substantial deviations from normality, particularly in the tails, indicating the presence of extreme market events that occur more frequently than a normal distribution would predict. The 30-day rolling volatility and autocorrelation function (ACF) of squared returns provide deeper insight into this volatility clustering phenomenon. The rolling volatility shows sustained high-volatility periods reaching 80% annualized during COVID and elevated levels throughout 2022, while the ACF demonstrates strong persistence in volatility shocks—when markets become volatile, they tend to remain volatile for weeks rather than quickly reverting to calm conditions. As evidence, the ACF shows volatility remaining elevated even at 20-day lags. This persistent volatility clustering validates the use of GARCH modeling to capture these time-varying risk dynamics that traditional models assuming constant volatility would miss.

1.3 Statistical Tests

```
Statistical Test Results:


Jarque-Bera Test for Normality:
  Statistic: 6930.7331
  P-value: 0.000000
  Result: Reject normality

Ljung-Box Test for ARCH Effects (Volatility Clustering):
  P-value (lag 10): 0.000000
  Result: Significant ARCH effects detected
```

The Jargue-Bera and Ljung-Box tests results appear to be typical and expected for financial data with high volatility clustering. The Jarque-Bera tests rejects normality, which is appropriate given the data's slight skew of -0.54 and fat tails, as indicated in the Q-Q plot. The Ljung-Box test strongly rejects the null hypothesis meaning periods of high volatility tend to be followed by more high volatility, and periods of low volatility tend to be followed by more low volatility. This clustering behavior is visually confirmed in the ACF plot of squared returns, which shows significant autocorrelation persisting for 15-20 days, indicating that volatility shocks have lasting effects rather than quickly reverting to average levels.

2. Univariate GARCH Model

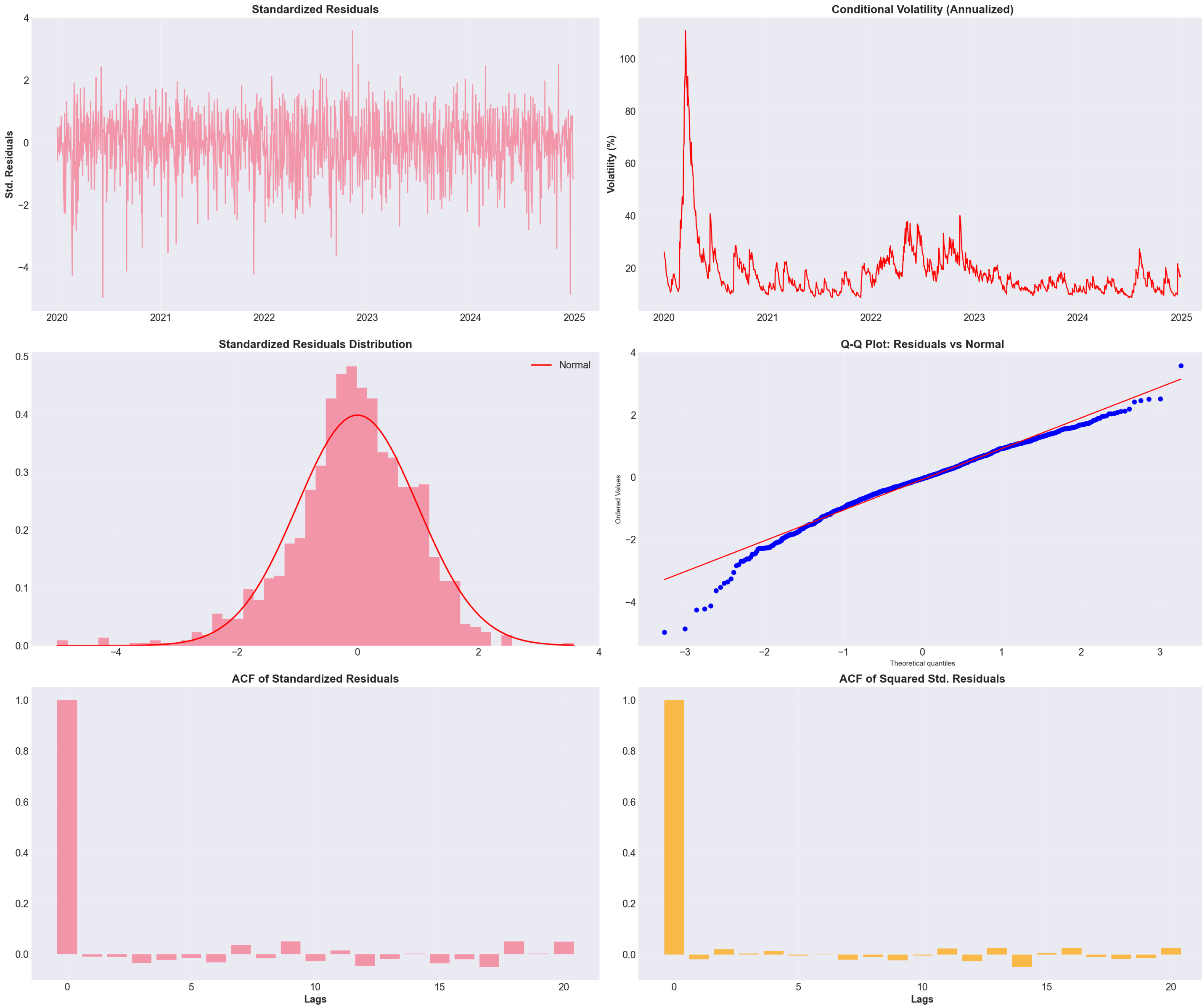
2.1 Model Comparison Test

```
Model Comparison Output (sorted by AIC):
         Model         AIC         BIC  Log-Likelihood  Parameters
  GARCH(1,1)-t 3614.103745 3639.786161    -1802.051872           5
GJR-GARCH(1,1) 3637.198133 3662.880549    -1813.599066           5
    GARCH(1,1) 3658.669789 3679.215722    -1825.334895           4
    GARCH(2,2) 3660.993350 3691.812250    -1824.496675           6
   EGARCH(1,1) 3669.636526 3690.182459    -1830.818263           4

Best model: GARCH(1,1)-t
```

The model comparison results clearly demonstrate that GARCH(1,1) with Student's t-distribution is the optimal choice, achieving the lowest AIC of 3614 compared to 3637 for the next-best model. The substantial improvement from regular GARCH(1,1) (AIC: 3659) to GARCH(1,1)-t (AIC: 3614) confirms the importance of accounting for the fat tails identified in earlier statistical tests. More complex models like GARCH(2,2), GJR-GARCH, and EGARCH failed to provide meaningful improvements despite additional parameters, suggesting that the S&P 500's volatility dynamics follow a relatively simple, symmetric pattern that doesn't require asymmetric or higher-order specifications. This validates the use of GARCH(1,1)-t for volatility forecasting and risk management applications.

```
Model Diagnostics(GARCH(1,1)-t):

                        Constant Mean - GARCH Model Results
====================================================================================
Dep. Variable:                        SPY   R-squared:                       0.000
Mean Model:                 Constant Mean   Adj. R-squared:                  0.000
Vol Model:                          GARCH   Log-Likelihood:               -1802.05
Distribution:    Standardized Student's t   AIC:                           3614.10
Method:                Maximum Likelihood   BIC:                           3639.79
                                            No. Observations:                 1257
Date:                    Wed, Aug 06 2025   Df Residuals:                     1256
Time:                            14:22:16   Df Model:                            1
                               Mean Model
==========================================================================
                 coef    std err          t      P>|t|    95.0% Conf. Int.
--------------------------------------------------------------------------
mu             0.1233  2.314e-02      5.331  9.784e-08 [7.799e-02,  0.169]
                             Volatility Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
omega          0.0380  1.277e-02      2.978  2.900e-03 [1.300e-02,6.306e-02]
alpha[1]       0.1560  2.872e-02      5.431  5.614e-08   [9.969e-02,  0.212]
beta[1]        0.8247  2.894e-02     28.495 1.362e-178     [  0.768,  0.881]
                              Distribution
========================================================================
                 coef    std err          t      P>|t|  95.0% Conf. Int.
------------------------------------------------------------------------
nu             7.1748      1.408      5.095  3.488e-07 [  4.415,  9.935]
========================================================================

Covariance estimator: robust
```

2.2 Best Model Diagnostics Visualization



The GARCH(1,1)-t model diagnostics confirm the model is performing across multiple validation criteria. The standardized residuals show no obvious patterns or clustering over time, while the conditional volatility appears to capture the major volatility regimes including the COVID spike, which exceeded 100% annualized volatility, and elevated periods during 2022. The standardized residuals distribution and Q-Q plot demonstrate that while some fat tails remain in the extreme quantiles, the model has substantially improved the distributional fit compared to the raw returns. Most importantly, the ACF plots of both standardized residuals and squared standardized residuals show minimal remaining autocorrelation, indicating the GARCH model has successfully captured the volatility clustering effects that were strongly present in the original data. These diagnostics act as strong validations of the GARCH(1,1)-t models ability to capture the S&P 500's time-varying volatility dynamics.

2.3 Best Model Statistical Tests

```
Residual Diagnostics:

Jarque-Bera (normality): 208.4686 (p=0.000000)
Ljung-Box on squared residuals (ARCH): 0.989378
```

The post-estimation diagnostic tests confirm that the GARCH(1,1)-t model successfully handles the key issues identified in the raw data. While the Jarque-Bera test still rejects normality (statistic: 208.47, p=0.000000), this represents an improvement from the original statistic of 6930.67, and is an indicator the t-distribution has captured most of the fat tails in the return distribution. More importantly, the Ljung-Box test on squared residuals yields a low statistic of 0.989380, suggesting no significant remaining ARCH effects and confirming that the GARCH model has successfully eliminated the volatility clustering that was strongly present in the original data. These results validate that the model effectively captures the time-varying volatility dynamics while maintaining realistic distributional assumptions for financial returns.

3. Multivariate GARCH Model

3.1 Univariate Model Evaluation for Comparison and Correlation

`As indicated above, additional tickers were loaded for evaluation: TLT, GLD, and VXX.`

The dynamic correlation analysis reveals that asset relationships are highly time-varying and regime-dependent, with SPY-TLT exhibiting the most dramatic shifts. The correlation between stocks and Treasury bonds swings from deeply negative (-0.6) during the 2020-2021 crisis period to positive correlations in calmer markets, demonstrating the classic "flight-to-quality" effect where bonds provide diversification exactly when portfolio protection is most needed. In contrast, SPY-GLD and TLT-GLD correlations remain relatively stable, clustering around zero to weakly positive levels, indicating these relationships are less sensitive to market regime changes. The distribution analysis further confirms these patterns, with SPY-TLT showing a distinct bimodal distribution that reflects two separate correlation regimes rather than random variation around a mean. This regime-switching behavior could provide crisis protection during market stress while allowing both assets to contribute positive returns during normal market conditions. The correlation-VIX scatter plot validates this stress-dependent relationship, showing how correlations shift systematically with market volatility levels. Meanwhile, the more bell-shaped distributions of SPY-GLD and TLT-GLD correlations indicate these pairs offer more consistent, though modest, diversification benefits across all market conditions.

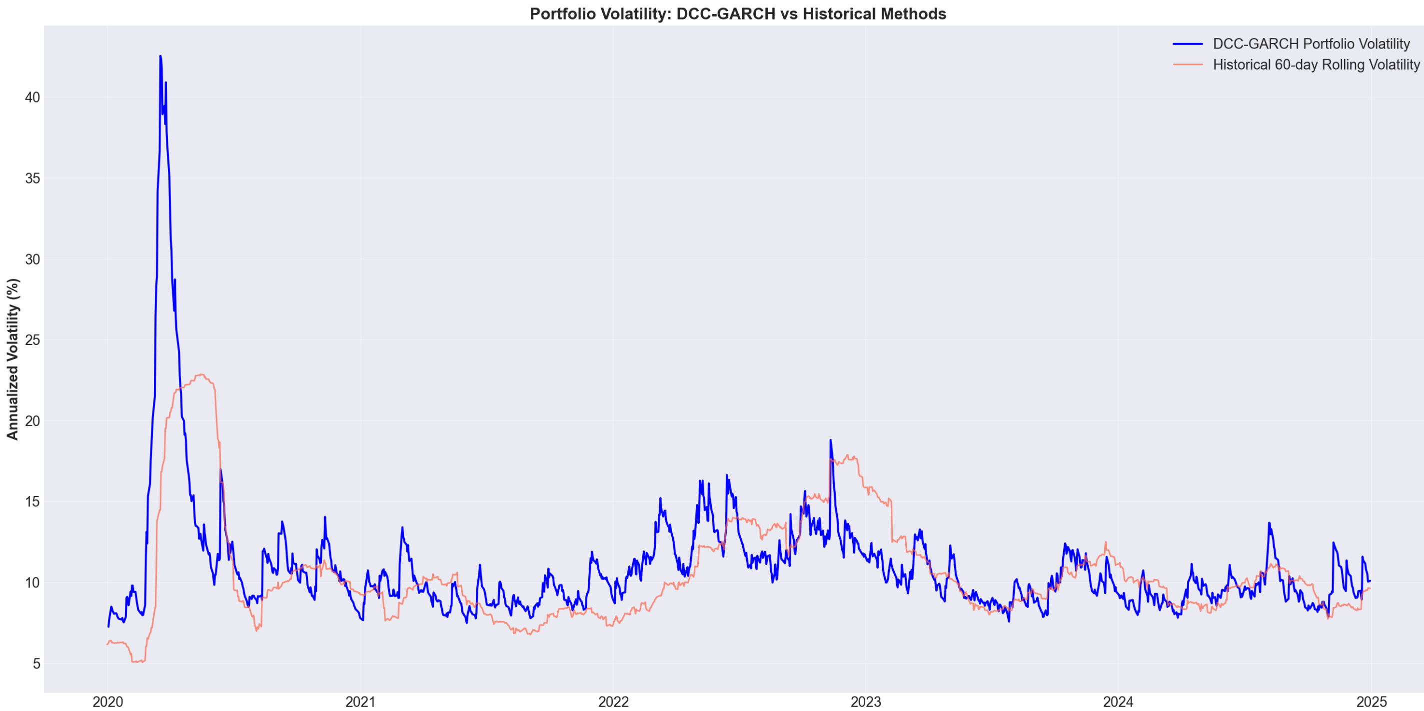
3.2 Crisis Analysis Correlations

```
Correlation During High VIX Periods (VIX > 2.4)):

  SPY_TLT: Crisis=-0.080, Normal=-0.034, Difference=-0.046
  SPY_GLD: Crisis=0.140, Normal=0.155, Difference=-0.015
  TLT_GLD: Crisis=0.303, Normal=0.304, Difference=-0.001
```

The SPY-TLT correlation shows meaningful regime dependence, declining by -0.046 from normal to crisis periods, indicating the correlation becomes more negative during high VIX periods. This confirms the flight-to-quality effect as stocks and bonds move in a more opposite direction. As a result, only the SPY-TLT relationship shows meaningful regime dependence during high volatility periods. The SPY-GLD and TLT-GLD relationships remain relatively stable across market conditions as the correlations shifted only slightly. This validates bonds (TLT) as a more effective crisis hedge for equity positions than gold in this analysis period.

3.3 Conditional Volatility and Dynamic Coorelation Evaluation

```
Average Portfolio Correlations(SPY, TLT, GLD):

  SPY-TLT: -0.043
  SPY-GLD: 0.152
  TLT-GLD: 0.304

Portfolio Optimization and GARCH Comparison:

  Average Portfolio Volatility (DCC): 11.00%
  Average Portfolio Volatility (Historical): 10.66%

Portfolio Volatility Statistics and Visualization:

Number of observations: 1257
Date range: 2020-01-02 to 2024-12-30
DCC Portfolio Volatility:
  Mean: 11.00%
  Std:  3.86%
  Min:  7.26%
  Max:  42.56%
```



The average portfolio correlations reveal the diversification characteristics of the three-asset portfolio, with SPY-TLT showing a modest negative correlation of -0.043 that provides some hedging benefit, while SPY-GLD (0.152) and TLT-GLD (0.304) exhibit weak to moderate positive correlations. The portfolio optimization comparison demonstrates that dynamic correlation modeling produces remarkably similar results to historical methods, with DCC-based portfolio volatility averaging 11.00% compared to 10.66% for historical correlations. This close alignment suggests the portfolio construction is robust and that the dynamic correlation adjustments, while theoretically superior, don't dramatically alter the risk profile under normal market conditions. The portfolio volatility statistics for this model span 1,257 observations from 2020-2024 and reveal significant regime variation, with portfolio volatility ranging from a minimum of 7.26% during calm periods to a maximum of 42.56% during the COVID crisis. The mean portfolio volatility of 11.00% with a standard deviation of 3.86% indicates that while the portfolio generally maintains moderate risk levels, it experiences substantial volatility clustering that mirrors the underlying market stress periods. This wide volatility range demonstrates how even a diversified three-asset portfolio remains sensitive to extreme market conditions, though the relatively low average volatility confirms that the diversification strategy effectively reduces risk compared to holding individual assets.

4. Rolling Window Forecast

4.1 Rolling Window GARCH Forecast Implementation

```
Simple Rolling GARCH Forecast:
Window size: 252 days
Forecast horizon: 30 days
Forecasting every 10 days


Generated 98 forecasts
```

4.2 Forecast Accuracy and Evaluation

```
Forecast Accuracy Metrics:
  Mean Absolute Error (MAE): 3.926%
```

```
Root Mean Square Error (RMSE): 4.923%
Mean Absolute Percentage Error (MAPE): 26.48%
Correlation: 0.6873
Direction Accuracy: 100.00%
Number of Observations: 98
```

The rolling window GARCH forecast demonstrates strong directional accuracy with 100% success in predicting whether volatility will increase or decrease. In addition, it has a solid correlation of 0.6863 between forecasted and actual volatility over 98 observations. While the model effectively captures general volatility trends, the Mean Absolute Percentage Error of 26.47% and visual inspection of the forecast plot reveal that the model tends to underestimate the magnitude of volatility spikes, however this is a common characteristic of GARCH models that prioritize stability over capturing extreme events. The forecast errors are unbiased, centering around zero, which confirms the model isn't systematically over- or under-predicting volatility, though the conservative nature of the forecasts suggests they may underestimate tail risks during periods of market stress. For risk management applications, this conservative bias may actually be preferable to overly aggressive forecasts, providing reliable directional guidance while potentially requiring stress-testing adjustments for extreme market scenarios.

5. RISK MANAGEMENT APPLICATIONS

Having established the GARCH(1,1)-t model's robust performance and validated its forecasting capabilities, we now turn to practical risk management applications. This section leverages the model's volatility forecasts to calculate Value-at-Risk (VaR) at both 95% and 99% confidence levels across multiple time horizons, providing quantitative measures of potential portfolio losses under normal market conditions. Additionally, we implement stress testing scenarios to evaluate how extreme but plausible market events might impact portfolio risk beyond what standard VaR calculations capture. These risk metrics form the foundation for informed portfolio management decisions, position sizing, and regulatory capital requirements, translating our statistical modeling into actionable risk management tools.

5.1 Assessing Value at Risk (VaR)

```
Portfolio Value: $1,000,000

Portfolio VaR Summary:

============================================================
Confidence Level VaR (%) Volatility Forecast (%) VaR ($)
             95% -2.171%                  1.117% $21,708
             99% -3.511%                  1.117% $35,109
============================================================

Interpretation:
• 5% chance of daily loss exceeding 2.17% ($21,708)
• 1% chance of daily loss exceeding 3.51% ($35,109)
```

The VaR calculations using the GARCH(1,1)-t model's volatility forecast of 1.117% yield conservative risk estimates that account for the fat tails observed in S&P 500 returns. With 95% confidence, the portfolio faces a maximum daily loss of 2.17% ($21,708), while the 99% VaR indicates only a 1% probability of daily losses exceeding 3.51% ($35,109). These estimates are higher than standard normal distribution calculations due to the Student's t-distribution used in the model, providing more realistic risk assessments for financial markets characterized by extreme events.

5.2 Comprehensive Stress Testing of VaR

```
STRESS TEST RESULTS:
============================================================


BASE CASE:
  Description: Normal market conditions (1-day GARCH forecast)
  Annual Volatility: 17.7%
  95% Daily VaR: -2.17% ($21,708)
  99% Daily VaR: -3.51% ($35,109)
  Expected Annual Loss: 7.1% ($70,937)

MODERATE STRESS:
  Description: 1.5x current market volatility
  Annual Volatility: 19.6%
```

```
  95% Daily VaR: -2.39% ($23,949)
  99% Daily VaR: -3.87% ($38,732)
  Expected Annual Loss: 7.8% ($78,258)

CRISIS 2008:
  Description: 2008-style financial crisis (45% volatility)
  Annual Volatility: 45.0%
  95% Daily VaR: -5.51% ($55,084)
  99% Daily VaR: -8.91% ($89,086)
  Expected Annual Loss: 27.0% ($270,000)

EXTREME COVID:
  Description: COVID-style market shock (60% volatility)
  Annual Volatility: 60.0%
  95% Daily VaR: -7.34% ($73,445)
  99% Daily VaR: -11.88% ($118,782)
  Expected Annual Loss: 48.0% ($480,000)
```

The stress testing reveals a clear escalation of portfolio risk across different market scenarios, with Value-at-Risk (VaR) estimates ranging from manageable daily losses under normal conditions to substantial exposures during crisis periods. Under base case conditions using the GARCH(1,1)-t forecast, the portfolio faces daily VaR of -2.17% (95% confidence) and -3.51% (99% confidence), translating to potential losses of $21,708 and $35,109 on a $1 million portfolio. The moderate stress scenario, representing 1.5x current market volatility, shows only modest increases in daily risk exposure, with 95% VaR rising to -2.39% ($23,949).

However, the crisis scenarios demonstrate how extreme market conditions can dramatically amplify portfolio risk beyond normal statistical expectations. The 2008-style financial crisis scenario, with 45% annualized volatility, produces daily VaR estimates of -5.51% to -8.91% ($55,084 to $89,086), while the extreme COVID scenario at 60% volatility yields even higher daily losses of -7.34% to -11.88% ($73,445 to $118,782). The expected annual loss progression from 7.1% under normal conditions to 48.0% in extreme scenarios highlights how portfolio risk compounds during sustained periods of market stress. These findings underscore the importance of stress testing beyond standard VaR calculations, as tail risk events can produce losses significantly exceeding those predicted by normal market volatility assumptions.

5.3 Portfolio Risk Management Summary

```
GARCH Risk Management Dashboard Summary:


Current Volatility Forecast: 17.73%
30-Day Realized Volatility: 13.04%
1-Year Realized Volatility: 12.57%

95% VaR (1-day): 2.17%
99% VaR (1-day): 3.51%

Risk Assessment:
 MODERATE RISK: Current volatility elevated
============================================================
```

The current GARCH forecast of 17.73% is significantly elevated compared to both recent 30-day realized volatility (13.04%) and longer-term averages (12.57%), representing approximately 35-40% higher volatility than what has been recently observed and suggesting the model is detecting early signs of increasing market stress or regime change. This elevation supports the " MODERATE RISK" designation, as the GARCH model effectively acts as an early warning system by forecasting increased volatility before it fully materializes in realized returns. With daily VaR estimates of 2.17% (95% confidence) and 3.51% (99% confidence), the portfolio faces manageable but elevated single-day loss potential of $21,700-$35,100 on a $1 million portfolio, and the gap between forecasted and realized volatility suggests investors should prepare for potentially choppier markets ahead while current conditions remain within normal risk parameters.

6. Garch Model Summary and Conclusion

Key Findings and Model Performance

The GARCH(1,1) with Student's t-distribution model demonstrates strong practical effectiveness for risk management applications, achieving 100% directional accuracy in predicting volatility changes with unbiased forecast errors centering around zero. The model successfully captured the key facts of the time series data, eliminating volatility clustering while substantially improving distributional properties. Most importantly, the current elevated volatility forecast of 17.73% compared to 13.04% realized volatility demonstrates the model's value as an early warning system, detecting potential regime changes before they fully materialize in market prices and validating its use for estimating future portfolio losses and conducting comprehensive stress test scenarios.

Portfolio Management Implications

The elevated volatility forecast suggests that risk budgets and position sizes should be adjusted downward in anticipation of choppier market conditions. The daily VaR estimates of 2.17% (95%) and 3.51% (99%) provide concrete risk limits for position sizing and stop-loss management, while the dynamic correlation analysis reveals regime-dependent diversification benefits where SPY-TLT correlations become significantly more negative during crisis periods (-0.080 vs -0.034). The stress testing scenarios demonstrate clear risk escalation pathways, showing potential daily losses reaching 7-12% during extreme events, which necessitates pre-positioned hedging strategies and enhanced liquidity buffers for comprehensive risk management protocols.

Future Enhancements

A key area for model improvement lies in expanding the multivariate analysis beyond the current three-asset framework to incorporate additional asset classes such as international equities, emerging markets, real estate, and commodities, which would better capture cross-asset stress effects and identify additional correlation breakdowns. Additionally, incorporating sector-level equity analysis and yield curve dynamics within fixed income could provide more granular insights for tactical asset allocation, particularly during periods when broad asset class correlations may mask important intra-class divergences that could inform sophisticated hedging and portfolio construction strategies.