University of Warsaw

Faculty of Economic Sciences

Financial Volatility Modeling and Forecasting with GARCH and DCC-GARCH

Financial Econometrics and Geoinformatics in Social Sciences using R

Giacomo Fantato · **K-16958**

g.fantato@student.uw.edu.pl

Contents

Introduction

General description

In recent years, understanding the dynamics of financial market volatility and the interdependence among global equity indices has become increasingly important for risk management and portfolio allocation. Traditional GARCH models have proven effective in capturing time-varying volatility in univariate time series; however, they fall short when analyzing co-movements across multiple assets. To address this limitation, multivariate GARCH models, particularly the Dynamic Conditional Correlation (DCC-GARCH) model proposed by Engle (2002)[1], have emerged as powerful tools for modeling time-varying correlations. This project aims to apply the DCC-GARCH framework to a selection of major international stock indices—namely, the S&P 500, NASDAQ 100, DAX, and Nikkei 225—to investigate how volatility and correlations evolve over time and how these relationships may reflect broader economic and market conditions.
The main motivation is understanding the interdependence between global financial markets is essential for investors, policymakers, and risk managers. The increased integration of economies and capital markets has amplified the need to monitor not only individual market volatility but also the dynamic relationships between markets. Sharp increases in market correlations during periods of financial stress can significantly reduce the benefits of diversification, making it crucial to study how these correlations evolve over time.

Overview of GARCH and DCC-GARCH Models

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, introduced by Bollerslev (1986), has become a cornerstone in modeling time-varying volatility in financial time series. It extends the ARCH model by allowing for a lagged structure in conditional variance, making it suitable for capturing volatility clustering—a common feature in asset returns. However, univariate GARCH models are insufficient for analyzing multiple assets jointly. The Dynamic Conditional Correlation (DCC-GARCH) model, developed by Engle (2002), addresses this limitation by modeling both individual volatilities and their evolving correlations. It provides a flexible and computationally feasible approach to studying multivariate volatility dynamics while preserving the advantages of univariate GARCH specifications.

[1]Engle, Robert. 2001. "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics." Journal of Economic Perspectives 15 (4): 157–168.

Data Collection, Preprocessing and Exploratory Data Analysis (EDA)

Visual inspection and data cleaning

Before computing the daily log returns based on closing prices for each index, I performed a series of preliminary checks to ensure data consistency and suitability for modeling. First, I examined each dataset for missing values. Observations containing `NA` values—typically due to market holidays or data unavailability—were removed to avoid issues during model estimation.

Next, I produced time series plots of the log returns for all four indices (S&P 500, NASDAQ 100, DAX, and Nikkei 225) to visually inspect their behavior over time. These plots revealed the presence of volatility clustering, a common feature in financial returns and a key motivation for using GARCH-type models.

Finally, I've computed summary statistics for each return series, including mean, standard deviation, minimum, and maximum values. These descriptive measures helped confirm the high-frequency characteristics of financial data, such as low average returns and heavy tails.

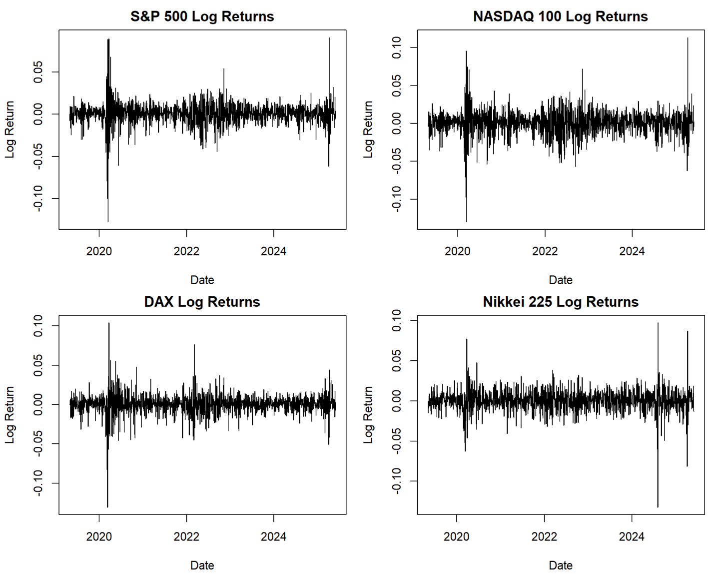


Figure 1: Log returns of major stock indices: S&P 500, NASDAQ 100, DAX, and Nikkei 225

Figure 1 displays the daily log returns of the S&P 500, NASDAQ 100, DAX, and Nikkei 225 indices from May 2019 to June 2025. All four time series exhibit periods of increased volatility followed by more stable phases, a phenomenon known as volatility clustering. Notably, large spikes in volatility are visible during major market events such as the COVID-19 pandemic in early 2020. The presence of such dynamics confirms the suitability of GARCH-type models, which are designed to capture time-varying volatility.

Time Series Plots of Price Levels and Returns: to visually inspect the behavior of the financial series, I have produced time series plots of both the price levels and the log returns for the four stock indices under study. The price level plots highlight the overall trends and long-term growth patterns of each index, while the log return plots reveal the short-term fluctuations and volatility characteristics of the series. In the

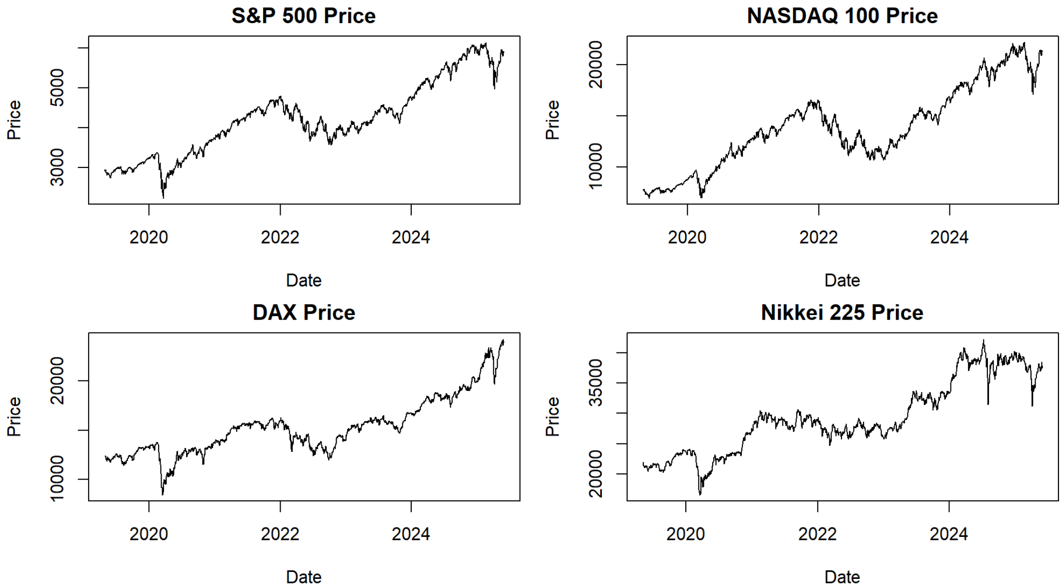


Figure 2: Daily closing prices of the S&P 500, NASDAQ 100, DAX, and Nikkei 225 indices from 2019 to 2025.

log return plots, I have observed a clear evidence of volatility clustering, where periods of high volatility are followed by similar periods. This behavior is consistent with the assumptions underlying GARCH-type models and further motivates their application in modeling and forecasting financial market volatility.

Summary Statistics of Log Returns

As a first step in the exploratory data analysis, I have computed the summary statistics for the daily log returns of the four selected stock indices: S&P 500, NASDAQ 100, DAX, and Nikkei 225. This preliminary analysis allows us to understand the general distributional properties of the returns, detect asymmetries, and evaluate the presence of heavy tails—typical characteristics of financial time series.

For each return series, I have computed the minimum, first quartile, median, mean, third quartile, and maximum, as well as the standard deviation, skewness, and kurtosis.

```
--- S&P 500 ---
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-0.1276522 -0.0049401  0.0008808  0.0004550  0.0069653  0.0908949
Standard Deviation: 0.01321853
Skewness: -0.6433418
Kurtosis: 17.41142

--- NASDAQ 100 ---
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.130032 -0.007008  0.001422  0.000659  0.009411  0.113528
Standard Deviation: 0.01617487
Skewness: -0.3389242
Kurtosis: 10.18706

--- DAX ---
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-0.1305486 -0.0048313  0.0009386  0.0004294  0.0069458  0.1041429
Standard Deviation: 0.01279954
Skewness: -0.7165735
Kurtosis: 16.24102

--- Nikkei 225 ---
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-0.1323408 -0.0063452  0.0008434  0.0003612  0.0072520  0.0973656
Standard Deviation: 0.01359203
Skewness: -0.4179862
Kurtosis: 14.42429
```

The S&P 500 returns show a slightly negative mean (0.0005), high kurtosis (17.41), and negative skewness (-0.64), indicating a distribution with heavier tails than the normal distribution and more frequent large negative returns. Similar features are observed in the DAX and Nikkei 225, both showing high kurtosis values of 16.24 and 14.42, respectively, and negative skewness. The NASDAQ 100 displays the highest standard deviation (0.0162), suggesting greater return volatility relative to the others, but it exhibits lower excess kurtosis (10.19) and skewness (-0.34) compared to the S&P 500 and DAX.
These statistics confirm that all return series are leptokurtic and left-skewed, justifying the use of models capable of capturing volatility clustering and heavy-tailed behavior, such as GARCH-type models.

Stationarity Tests

To verify the stationarity of the log-return series, I applied two complementary statistical tests: the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. The ADF test examines the null hypothesis of a unit root (non-stationarity), while the KPSS test assumes stationarity under the null hypothesis.

```
--- S&P 500 ---
ADF Test:
data:  x
Dickey-Fuller = -10.762, Lag order = 11, p-value = 0.01
alternative hypothesis: stationary

KPSS Test:
data:  x
KPSS Level = 0.040793, Truncation lag parameter = 7, p-value = 0.1

--- NASDAQ 100 ---
ADF Test:
data:  x
Dickey-Fuller = -10.691, Lag order = 11, p-value = 0.01
alternative hypothesis: stationary

KPSS Test:
data:  x
KPSS Level = 0.069178, Truncation lag parameter = 7, p-value = 0.1

--- DAX ---
ADF Test:
data:  x
Dickey-Fuller = -10.643, Lag order = 11, p-value = 0.01
alternative hypothesis: stationary

KPSS Test:
data:  x
KPSS Level = 0.10134, Truncation lag parameter = 7, p-value = 0.1

--- Nikkei 225 ---
ADF Test:
data:  x
Dickey-Fuller = -11.44, Lag order = 11, p-value = 0.01
alternative hypothesis: stationary
```

```
KPSS Test:
data:  x
KPSS Level = 0.029535, Truncation lag parameter = 7, p-value = 0.1
```

For all four indices—S&P 500, NASDAQ 100, DAX, and Nikkei 225—the ADF test strongly rejects the null hypothesis of a unit root, with test statistics well below the critical values and p-values less than 0.01. Simultaneously, the KPSS test fails to reject the null hypothesis of stationarity, with all p-values greater than 0.1.
Taken together, the results from both tests consistently indicate that the daily log returns of all four indices can be considered stationary. This finding is essential, as stationarity is a key assumption in GARCH-type models and necessary for meaningful volatility modeling and forecasting.

Volatility Clustering and ARCH Effects

To investigate the presence of volatility clustering—a common feature in financial time series, I've analyzed the autocorrelation of squared returns for each index. Figure 3 shows the autocorrelation functions (ACF) of squared log returns for the S&P 500, NASDAQ 100, DAX, and Nikkei 225. In all four series, the ACF plots of squared returns show

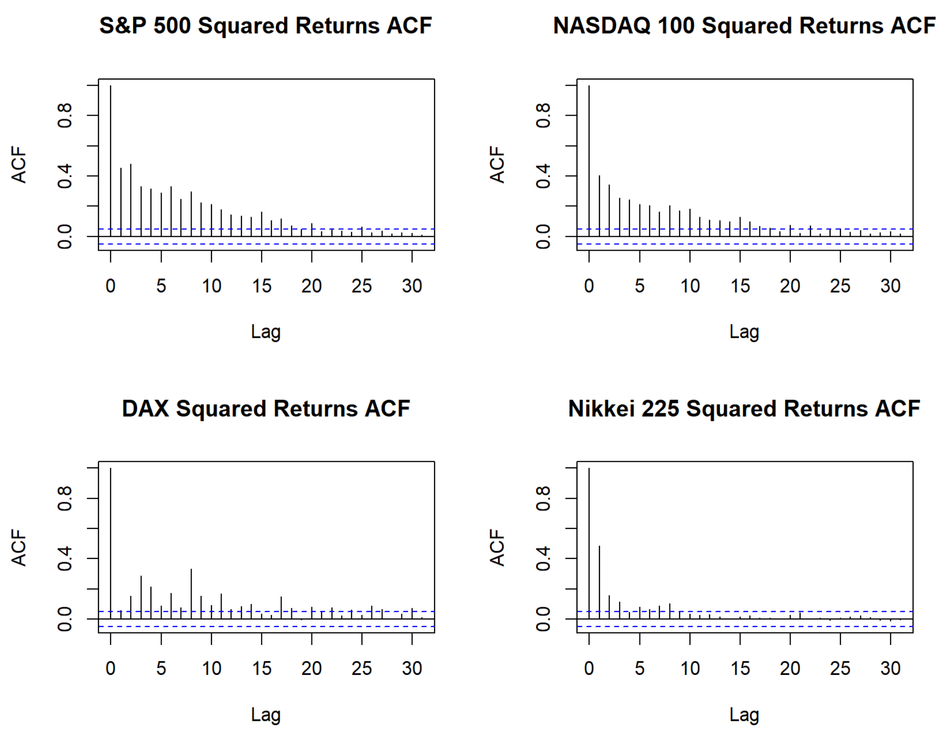


Figure 3: Autocorrelation of squared returns for S&P 500, NASDAQ 100, DAX, and Nikkei 225. Significant lags indicate the presence of volatility clustering.

significant autocorrelation, especially at lower lags, suggesting the presence of persistent conditional heteroskedasticity; this motivates the application of ARCH-type models.

To formally test for ARCH effects, could be conducted the ARCH LM-test with 10 lags. The results are strongly significant for all indices:

- S&P 500: $\chi^2 = 497.34$, df $= 10$, $p < 2.2 \times 10^{-16}$
- NASDAQ 100: $\chi^2 = 344.12$, df $= 10$, $p < 2.2 \times 10^{-16}$
- DAX: $\chi^2 = 335.24$, df $= 10$, $p < 2.2 \times 10^{-16}$
- Nikkei 225: $\chi^2 = 404.08$, df $= 10$, $p < 2.2 \times 10^{-16}$

In all cases, the null hypothesis of no ARCH effects is rejected at any reasonable significance level. These findings confirm the existence of time-varying volatility and validate the use of GARCH-type models for further analysis.

GARCH Model Estimation

Univariate GARCH(1,1) Model Estimation

I've estimated univariate GARCH(1,1) models for each of the four return series using the `rugarch` package in R. The specification includes a constant mean and a standard GARCH process for the conditional variance. The estimation was conducted under the assumption of normally distributed residuals.
The fitted models provided statistically significant GARCH parameters for all indices. In each case, the α_1 and β_1 coefficients were both significantly different from zero, indicating a strong presence of conditional heteroskedasticity and persistence in volatility.

To evaluate the adequacy of the model, several diagnostic tests could be analyzed:

- The **Ljung–Box tests** on both standardized residuals and squared residuals failed to reject the null hypothesis of no serial correlation in all cases. This indicates that the GARCH(1,1) models adequately captured the autocorrelation structure of the data.

- The **ARCH LM tests** at multiple lags produced non-significant results (p-values above 0.1), confirming that no significant ARCH effects remained in the standardized residuals. This suggests that the GARCH(1,1) specification was sufficient to model the time-varying volatility.

- The **Nyblom stability test** indicated stable parameters for most indices, except for DAX, where the joint statistic exceeded the 1% critical value. This suggests some parameter instability, which may warrant further investigation or model refinement.

- The **sign bias tests** were generally non-significant, with the exception of the S&P 500 and Nikkei 225. For these indices, the joint effect test indicated some asymmetric behavior in volatility responses to positive and negative shocks, which may motivate the use of asymmetric GARCH models such as GJR-GARCH or EGARCH in future work.

- The **Adjusted Pearson Goodness-of-Fit** test returned significant results across all indices, indicating a potential mismatch between the assumed normal distribution and the actual distribution of standardized residuals. This is consistent with the known heavy tails in financial return distributions.

In summary, the GARCH(1,1) models captured the key dynamics of volatility in each return series, with good residual diagnostics and no remaining ARCH effects. Nonetheless, some departures from model assumptions—such as non-normality and mild asymmetry—suggest that alternative specifications may provide further improvements in certain cases.

Conditional standard deviation

Figure 4 displays the estimated conditional standard deviations for the four indices, as obtained from the GARCH(1,1) models. The plots show the return series with ± 2 conditional standard deviations superimposed in red, providing a visual representation of time-varying volatility.

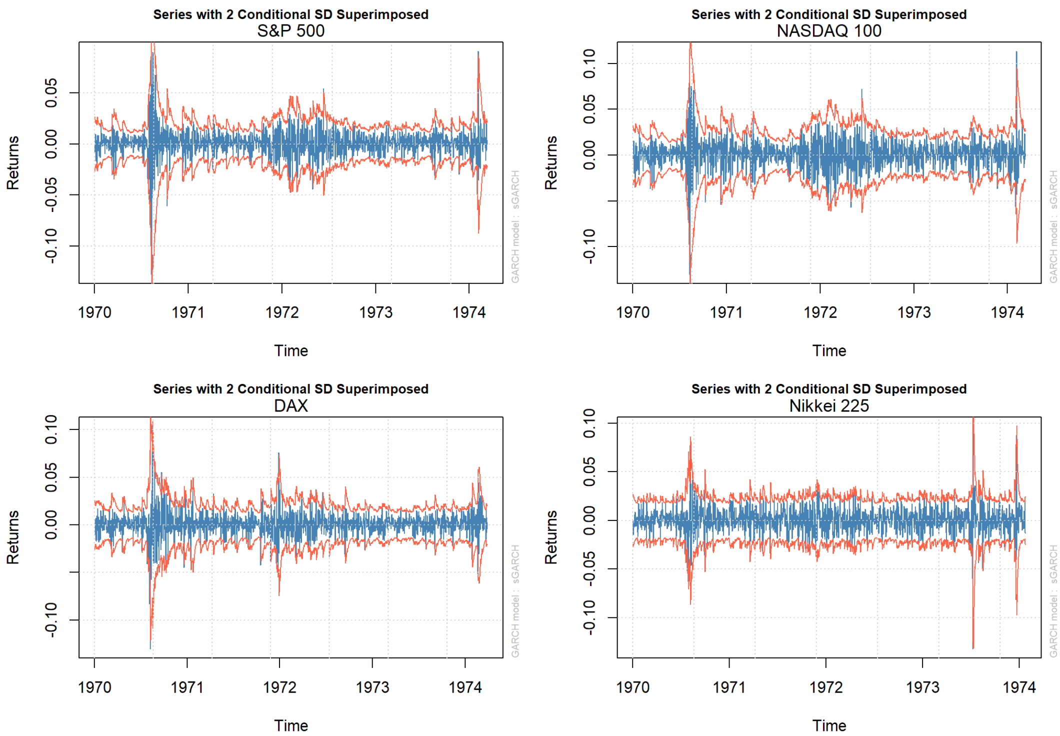


Figure 4: Conditional standard deviations from GARCH(1,1) models for S&P 500, NASDAQ 100, DAX, and Nikkei 225.

The following conclusions can be drawn:

- All four indices exhibit clear volatility clustering: periods of high volatility are followed by similar periods of high variation, and likewise for low volatility.

- The models successfully capture large spikes in conditional standard deviations during market shocks, especially around the early part of the series, consistent with financial crisis events.

- Volatility tends to decay over time after shocks, consistent with GARCH model assumptions.

- Despite overall good performance, the amplitude of conditional standard deviations varies by market. The NASDAQ 100 and Nikkei 225 show wider confidence bands in several periods, reflecting higher market sensitivity.
- The GARCH(1,1) models appear adequate in capturing dynamic volatility, making them suitable for forecasting and as input for multivariate extensions like DCC-GARCH.

Standardized Residuals Analysis

To assess the adequacy of the GARCH(1,1) model for each index, I examined the standardized residuals obtained from the model fits. Figure 5 shows the time series plots of standardized residuals for the S&P 500, NASDAQ 100, DAX, and Nikkei 225 indices.

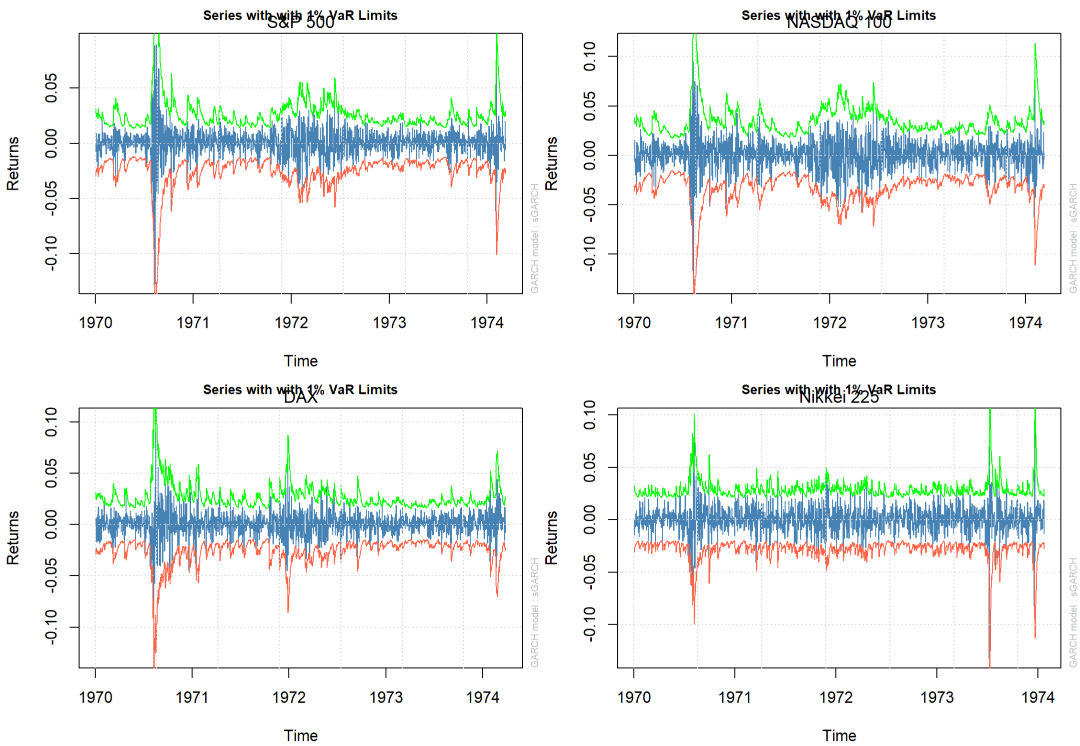


Figure 5: Standardized residuals from GARCH(1,1) models for S&P 500, NASDAQ 100, DAX, and Nikkei 225.

These residuals, which should ideally resemble white noise, allow us to verify whether the model has successfully captured the dynamics of conditional heteroskedasticity.

The standardized residual series do not exhibit clear patterns or autocorrelation, suggesting that the GARCH(1,1) model has effectively filtered out the conditional volatility component from the original return series. Additionally, there are no large clusters or persistent periods of high residual variance, which supports the model's ability to stabilize the variance over time. The residuals appear symmetric and randomly distributed, indicating no obvious signs of model misspecification. However, slight deviations from perfect whiteness, especially in the DAX and Nikkei 225 indices, may suggest that further improvements could be achieved by adopting more flexible specifications, such as GARCH models with Student-t innovations or asymmetric forms like GJR-GARCH or EGARCH. Overall, the GARCH(1,1) specification seems to have captured the conditional variance dynamics reasonably well for each of the four indices, as confirmed visually through the behavior of the standardized residuals.

Autocorrelation of Squared Residuals

To further evaluate the adequacy of the GARCH(1,1) models, could be examine the autocorrelation functions (ACF) of the squared standardized residuals for each index, as shown in Figure 6. These plots help detect any remaining conditional heteroskedasticity that may not have been captured by the model. In a well-specified GARCH model, the squared residuals should not display significant autocorrelation at any lag, implying that the volatility dynamics have been adequately modeled. The visual inspection of the ACF plots shows that autocorrelations fall within the confidence bounds across all indices, indicating the absence of strong ARCH effects in the residuals and confirming the appropriateness of the GARCH(1,1) specification for capturing time-varying volatility.

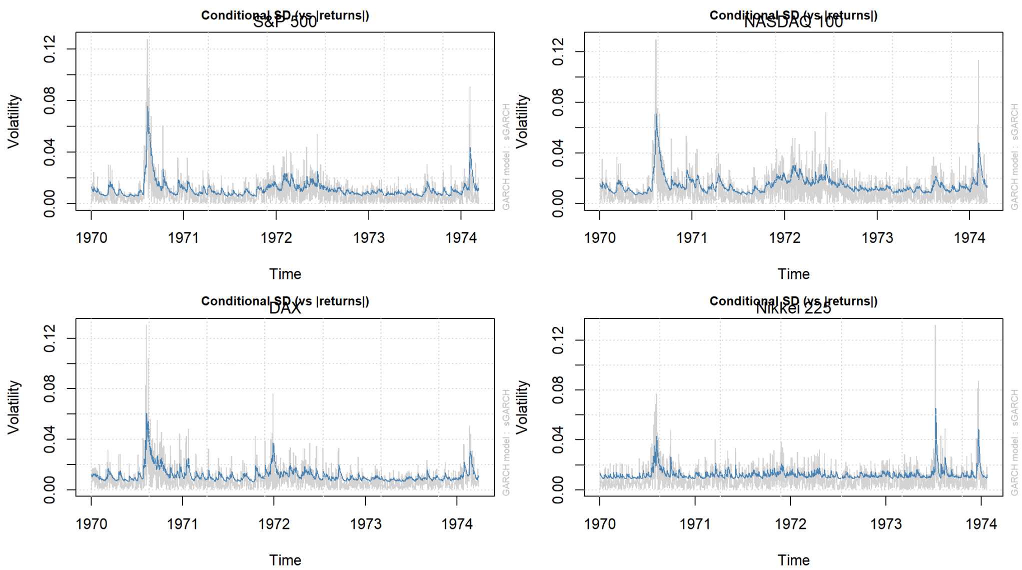


Figure 6: Autocorrelation of squared standardized residuals from GARCH(1,1) models for the four indices.

Q-Q Plot of Standardized Residuals

To evaluate the adequacy of the distributional assumptions underlying the GARCH(1,1) models, quantile-quantile (Q-Q) plots of the standardized residuals were examined for each index, as presented in Figure 7. These plots facilitate the assessment of deviations from normality and help validate the model's assumption of conditionally normally distributed errors. If the residuals follow a normal distribution, the points should align closely with the 45-degree reference line. The Q-Q plots reveal deviations from normality in the tails for all four indices, particularly in the upper and lower extremes.

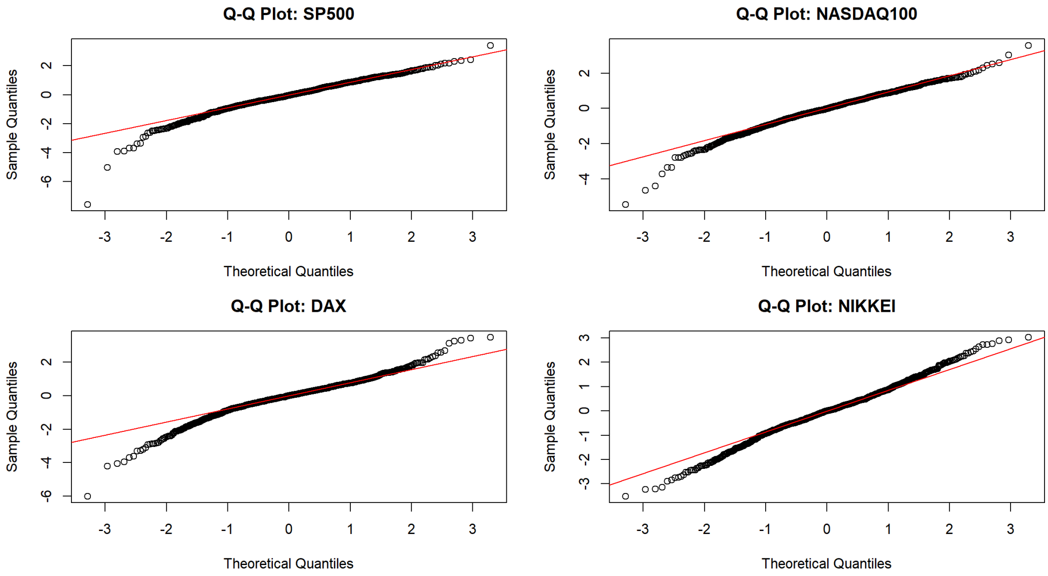


Figure 7: Q-Q plots of standardized residuals from GARCH(1,1) models for S&P 500, NASDAQ 100, DAX, and Nikkei 225.

This indicates that the standardized residuals exhibit heavier tails than expected under the normal distribution assumption. Such behavior is typical in financial return series and suggests that, although the GARCH(1,1) model captures the time-varying volatility well, the assumption of normally distributed errors may not be fully adequate. Alternative specifications, such as models incorporating Student-t or skewed distributions, could potentially provide a better fit for capturing tail risk.

Model Adequacy Testing via Residual Diagnostics

To assess the adequacy of the fitted GARCH(1,1) models, I'm going to conduct a residual diagnostics based on statistical tests. Specifically, i have applied the Ljung–Box test to both the standardized residuals and their squared values. The rationale is that if a GARCH model has successfully captured the conditional heteroskedasticity present in the return series, then the standardized residuals should resemble white noise, and no additional autocorrelation or volatility clustering should remain.

The Ljung–Box test on the standardized residuals tests the null hypothesis of no autocorrelation up to a specified lag (in this case, lag 20). A p-value greater than 0.05 suggests that residuals are not autocorrelated, which is desirable.

Similarly, the Ljung–Box test on the squared residuals tests for remaining ARCH effects. If the p-value is above 0.05, it implies that the model has sufficiently captured the time-varying volatility, and no conditional heteroskedasticity remains in the data.

```
test_garch_adequacy <- function(fit, name) {
  cat("\n---", name, "---\n")
  resid <- residuals(fit, standardize = TRUE)
  # Ljung-Box test on residuals
  test_resid <- Box.test(resid, lag = 20, type = "Ljung-Box")
```

```
  # Ljung-Box test on squared residuals
  test_resid_sq <- Box.test(resid^2, lag = 20, type = "Ljung-Box")
  cat("Ljung-Box on residuals (autocorrelation):\n")
  print(test_resid)
  cat("Ljung-Box on squared residuals (ARCH effect):\n")
  print(test_resid_sq)
  # Decision logic
  if (test_resid$p.value > 0.05 && test_resid_sq$p.value > 0.05) {
    cat("GARCH(1,1) appears adequate for", name, "\n")
  } else {
    cat("Consider extending the model for", name, ":\n")
    if (test_resid$p.value <= 0.05) cat("  → Residuals show
    autocorrelation.\n")
    if (test_resid_sq$p.value <= 0.05) cat(" → ARCH effects remain in
    squared residuals.\n")
    cat(" → Suggest trying GARCH(p,q), GJR-GARCH, or t-distribution\n")
  }
}
```

The automated testing function outputs these diagnostics and applies a decision rule:

- If both p-values exceed the 0.05 threshold, the GARCH(1,1) model is deemed adequate.
- If either p-value is below 0.05, this indicates model inadequacy—either in filtering autocorrelation or in removing ARCH effects—and more flexible models such as higher-order GARCH, GJR-GARCH, or models assuming non-normal innovations (e.g., Student-t distribution) may be more appropriate.

This procedure provides a structured and replicable method for validating the chosen GARCH specification across different financial time series (all the time series included in our analysis).

DCC-GARCH Model Estimation

In order to analyze the time-varying dependencies between the selected financial indices, I have estimated a multivariate volatility model known as the Dynamic Conditional Correlation GARCH (DCC-GARCH) model. This model is particularly well-suited for capturing the evolving correlation structure across multiple time series while accounting for individual volatility dynamics.

Theoretical background

The DCC-GARCH model, introduced by Engle (2002)[2], extends the univariate GARCH framework to a multivariate setting. It assumes that the conditional covariance matrix of asset returns can be decomposed into two components:

$$H_t = D_t R_t D_t$$

where:

- H_t is the conditional covariance matrix of asset returns at time t,
- D_t is a diagonal matrix of time-varying standard deviations (obtained from univariate GARCH models for each series),
- R_t is the time-varying conditional correlation matrix.

This separation allows for a flexible specification in which individual volatilities are modeled using standard univariate GARCH(1,1) processes, while the dynamic correlation structure is captured through a separate DCC process:

$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha \epsilon_{t-1} \epsilon_{t-1}^{\top} + \beta Q_{t-1}$$

$$R_t = \mathrm{diag}(Q_t)^{-1/2} Q_t \mathrm{diag}(Q_t)^{-1/2}$$

where Q_t is the covariance matrix of standardized residuals and $\bar{Q}$ is its unconditional expectation. The parameters α and β govern the persistence and responsiveness of the conditional correlation dynamics.

Estimation procedure

I proceed with the following steps:

1. The daily log returns of the S&P 500, NASDAQ 100, DAX, and Nikkei 225 indices are merged into a multivariate time series object.
2. Each series is individually modeled with a GARCH(1,1) specification under the assumption of conditional normality.
3. A DCC(1,1) model is estimated using the `rmgarch` package, which jointly estimates the univariate GARCH and the dynamic correlation process.
4. The time-varying pairwise correlations are extracted and plotted to visualize the evolution of cross-market dependencies.

[2]Engle, R. (2002). *Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models.* Journal of Business & Economic Statistics, 20(3), 339–350.

The main motivation is that modeling time-varying correlations is essential for understanding financial contagion, risk spillovers, and portfolio diversification opportunities. The DCC-GARCH model allows us to detect how correlations intensify or weaken during periods of market stress or tranquility, providing insights beyond what constant correlation models can offer.

Model output

Figure 10 displays the estimated dynamic conditional correlations between all pairwise combinations of the four selected indices: S&P 500, NASDAQ 100, DAX, and Nikkei 225. These correlations were obtained from the DCC-GARCH(1,1) model, which allows the conditional correlation matrix to evolve over time based on past standardized residuals.

The correlation between the S&P 500 and NASDAQ 100 remains consistently high, ranging between 0.85 and 0.95, highlighting the strong integration within the U.S. equity markets. Correlations between U.S. indices and the DAX also show relatively high but more variable levels, reflecting close transatlantic market connections. In contrast, correlations involving the Nikkei 225 are generally lower and more volatile, sometimes even becoming negative. This suggests that the Japanese market exhibits more independent dynamics relative to the Western markets.

Overall, the plots emphasize the importance of modeling correlations as time-varying, especially during periods of financial stress. Static correlation models would fail to capture such dynamic co-movement behavior, while the DCC framework effectively reveals the evolution and asymmetry in international market dependencies.

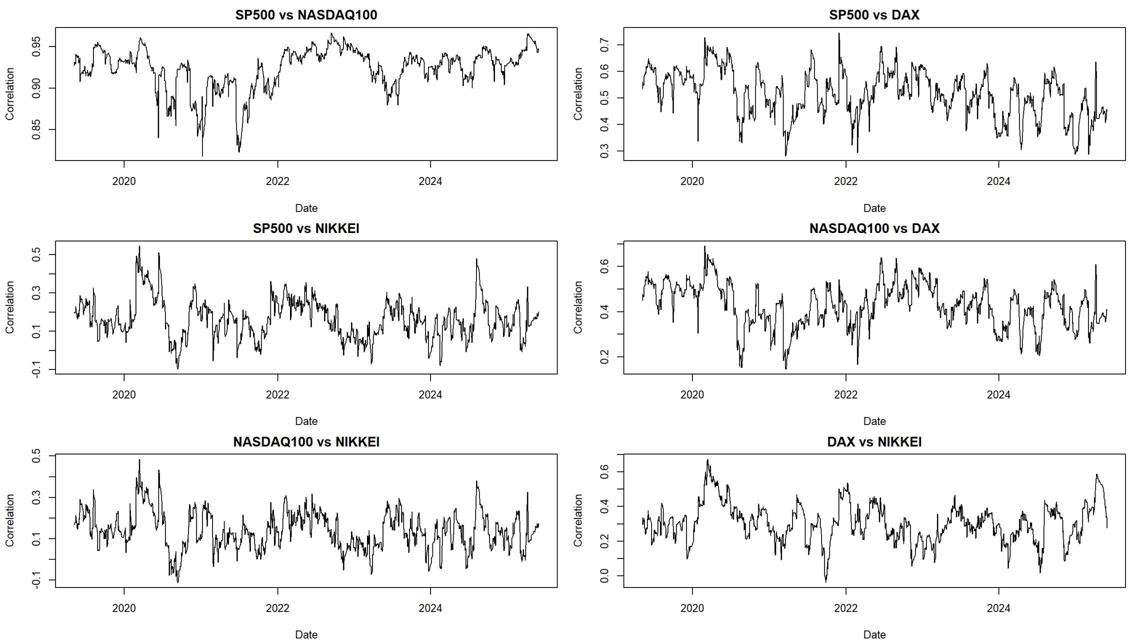


Figure 8: Pairwise dynamic conditional correlations estimated from the DCC-GARCH(1,1) model.

Analysis and interpretation

Empirical interpretation

The dynamic conditional correlations estimated through the DCC-GARCH(1,1) model reveal meaningful patterns in the co-movement of major global equity indices. Notably, the S&P 500 and NASDAQ 100 exhibit persistently high correlations across the entire sample period, often exceeding 0.9. This reflects the strong structural link between these two U.S. markets, which share similar economic drivers and investor bases.

In contrast, correlations between the U.S. and European markets (e.g., S&P 500 and DAX, NASDAQ 100 and DAX) are generally high but more volatile. These correlations tend to increase during periods of global financial stress—such as the COVID-19 outbreak in early 2020 or inflation shocks during 2022—indicating rising global market integration in times of uncertainty. This phenomenon, often referred to as "correlation breakdown" or "contagion," reduces the effectiveness of international diversification when it is most needed.

The Japanese market, represented by the Nikkei 225, shows weaker and more unstable correlations with Western markets. The lower average correlation with the S&P 500 and DAX suggests a degree of segmentation and potential for diversification benefits. However, periods of rising correlation can still be observed, such as during synchronized global monetary policy responses or shared macroeconomic shocks.

Spikes in correlation across indices often align with major events such as:

- **March 2020:** Global equity markets collapsed in response to the onset of the COVID-19 pandemic, triggering simultaneous declines and heightened co-movement across all indices.
- **February–March 2022:** The Russian invasion of Ukraine introduced significant geopolitical uncertainty, especially impacting European markets, which led to increased correlation with U.S. indices.
- **Throughout 2023:** Episodes of synchronized central bank tightening (e.g., by the Federal Reserve and ECB) led to heightened cross-market sensitivity to macroeconomic announcements.

Implications for portfolio management: while global diversification remains a valuable risk management strategy, the observed correlation dynamics demonstrate that asset co-movements are not constant and tend to increase during turbulent periods. This correlation clustering effect undermines the benefits of diversification exactly when investors seek protection. Therefore, portfolio strategies must account for time-varying correlation structures. The DCC-GARCH framework provides a more realistic tool for portfolio optimization, risk parity strategies, and dynamic hedging compared to static covariance assumptions. Investors should consider adjusting their allocation based on current correlation regimes rather than relying on long-term averages.

Rolling analysis of correlation stability

To assess the temporal stability of the correlations estimated by the DCC-GARCH model, I computed rolling statistics over a moving window of 90 trading days. For each pair of indices, I extracted the time-varying conditional correlations and then calculated the rolling mean and standard deviation to capture potential structural shifts or episodes of increased co-movement.

This step is crucial in evaluating the reliability of the DCC-GARCH specification for dynamic risk modeling. While the DCC framework allows correlations to evolve over time, examining their rolling behavior provides an additional layer of diagnostic insight. Periods of sustained high correlations or large fluctuations may indicate market stress, contagion, or shifts in investor sentiment that cannot be easily captured by static models.

Figure 9: Rolling 90-day average correlations with ± 1 standard deviation bands (in red) for each index pair. This plot highlights changes in correlation regimes and the stability of co-movements among international equity markets.

The resulting plots show, for each index pair, the trajectory of the 90-day rolling average correlation, accompanied by a red shaded band representing one standard deviation above and below the mean. The red band highlights periods of increased correlation volatility, which are of particular interest in risk management and asset allocation. Monitoring these dynamics is essential to understanding diversification effectiveness and to anticipating changes in the systemic integration of global markets.

GARCH(1,1) - VaR

The figure displays the in-sample Value-at-Risk (VaR) at the 5% level estimated using a separate GARCH(1,1) model for each of the four major equity indices: SP500, NASDAQ100, DAX, and NIKKEI. The gray lines represent the log-returns, the red lines show the estimated 5% VaR, and the blue dots indicate VaR violations—i.e., returns that fall below the predicted threshold.

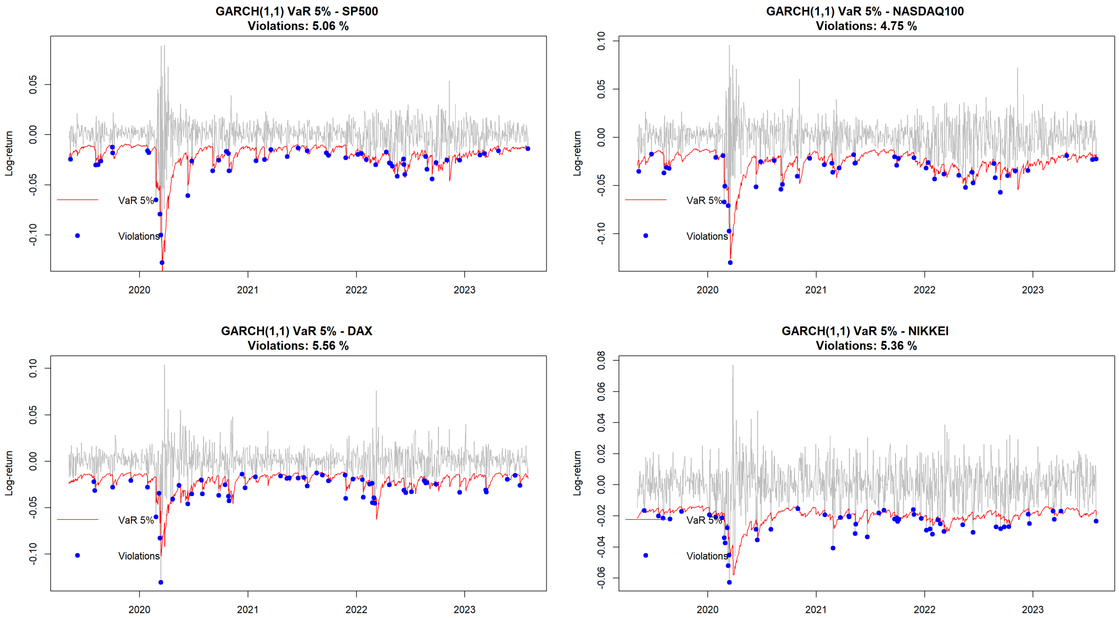


Figure 10: Pairwise dynamic conditional correlations estimated from the DCC-GARCH(1,1) model.

- **SP500:** The observed violation rate is approximately 5.06%, which is very close to the nominal 5% level. This suggests that the GARCH(1,1) model provides a well-calibrated risk estimate for the SP500 over the sample period.
- **NASDAQ100:** The violation frequency is slightly below the nominal level at 4.75%. This implies a marginally conservative VaR estimate—i.e., the model tends to slightly overestimate the downside risk.
- **DAX:** The violation rate is 5.56%, slightly higher than 5%. This may indicate a modest underestimation of tail risk, suggesting that the model might not fully capture volatility clustering in this index.
- **NIKKEI:** With a violation rate of 5.36%, the model performs reasonably well, though slightly under the expected confidence level. The fit appears robust but might benefit from heavier-tailed distributions.

Overall, the GARCH(1,1) model appears to be adequate for estimating VaR across all indices. The empirical violation rates are close to the theoretical 5% benchmark, with small deviations that could potentially be improved by incorporating alternative error distributions (e.g., Student-t) or more flexible GARCH structures.

DCC-GARCH(1,1) - VaR

To assess financial risk across major equity indices, I have estimated the 5% in-sample Value-at-Risk (VaR) using both univariate GARCH(1,1) and multivariate DCC-GARCH(1,1) models. While the GARCH(1,1) model is applied separately to each return series, capturing individual volatility dynamics effectively—with violation rates close to the nominal 5% level—it ignores interdependencies between assets. In contrast, the DCC-GARCH(1,1) model jointly estimates both conditional variances and time-varying correlations, offering a more comprehensive framework for capturing market co-movements. As shown in Figure 11, the DCC-GARCH model yields violation rates for all indices (SP500, NASDAQ100, DAX, and NIKKEI) ranging between 5.15% and 5.48%, confirming that the model is well-calibrated. Moreover, it accounts for periods of market stress and contagion, as evidenced by the synchronized volatility spikes during the 2020 COVID-19 shock. Given these properties, DCC-GARCH is especially suitable for risk management at the portfolio level. Therefore, although both models perform adequately in-sample, I recommend using the **DCC-GARCH(1,1)** model for out-of-sample forecasting, particularly in settings where asset interdependence plays a critical role.

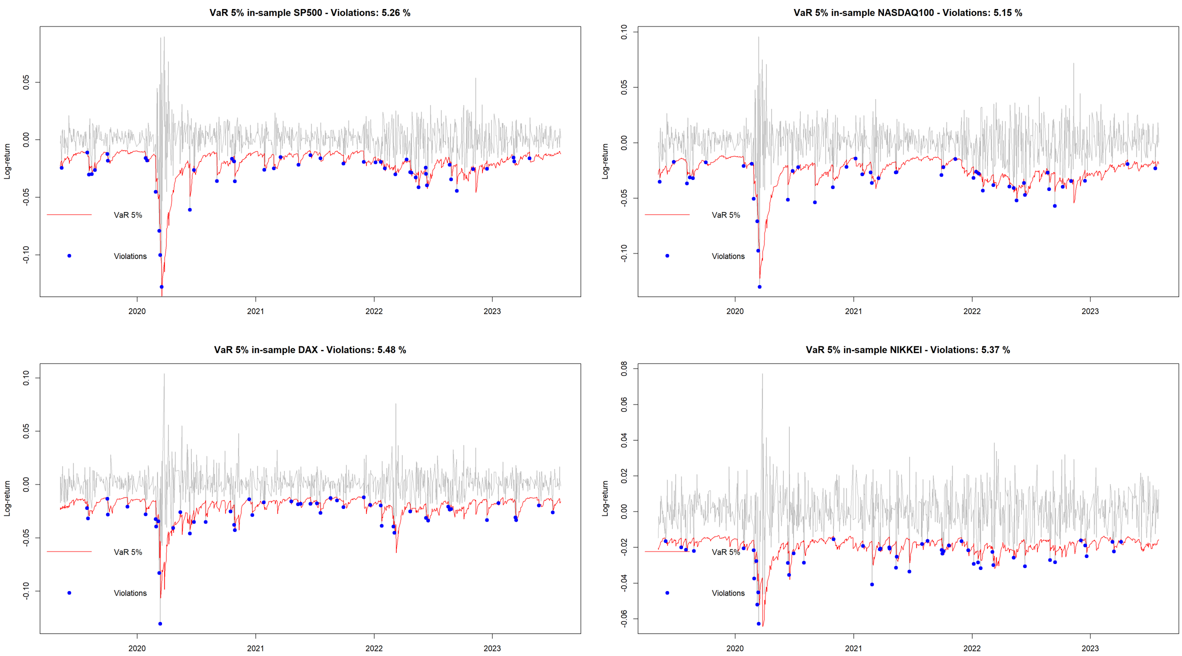


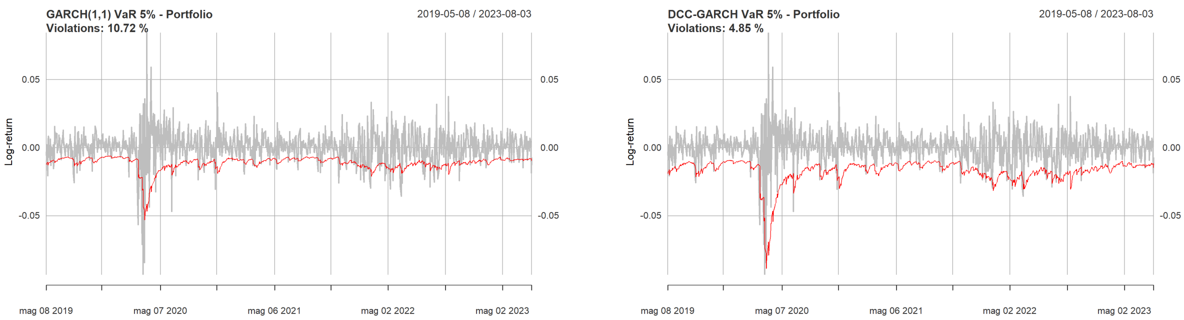
Figure 11: In-sample 5% VaR estimated using DCC-GARCH(1,1) for SP500, NASDAQ100, DAX, and NIKKEI. Gray lines show log-returns, red lines indicate the VaR threshold, and blue dots mark violations.

DCC-GARCH(1,1) vs GARCH(1,1) - equally weighted portfolio

Comparison of GARCH(1,1) and DCC-GARCH VaR on an Equally Weighted Portfolio. To evaluate the adequacy of volatility modeling for portfolio-level risk estimation, I'm going to compare the in-sample 5% Value-at-Risk (VaR) performance of two models: a naive approach based on univariate GARCH(1,1) models (ignoring asset correlations), and a multivariate DCC-GARCH(1,1) model that dynamically accounts for time-varying covariances among assets. Both models are applied to an equally weighted portfolio (25% allocation to each of SP500, NASDAQ100, DAX, and NIKKEI), over the period from May 2019 to August 2023.

As shown in Figure 12, the GARCH(1,1)-based portfolio VaR exhibits a violation rate of **10.72%**, which is more than double the theoretical level of 5%. This significant underestimation of downside risk stems from the fact that the model treats assets independently, thereby ignoring periods of heightened correlation, especially during market turbulence. In contrast, the DCC-GARCH(1,1) model yields a much more accurate violation rate of **4.85%**, which closely matches the expected 5% level.

The DCC-GARCH's superior performance arises from its ability to jointly estimate individual volatilities and dynamic conditional correlations, capturing co-movement patterns and contagion effects. This enables the model to produce tighter, more responsive VaR estimates in volatile market conditions. Therefore, based on the in-sample backtesting metric—the violation rate—**the DCC-GARCH(1,1) model clearly outperforms the univariate approach**, and should be preferred when managing portfolio risk.



GARCH(1,1) VaR – Violations: 10.72%

DCC-GARCH(1,1) VaR – Violations: 4.85%

Figure 12: In-sample 5% VaR estimation for an equally weighted portfolio using GARCH(1,1) vs. DCC-GARCH(1,1). The DCC model produces better calibrated risk estimates.

Forecast on portfolio and conclusion

In the rolling forecast procedure, a moving window of 300 observations was employed at each iteration. This window size was selected primarily due to computational constraints, as the available hardware resources did not permit the estimation of larger windows within reasonable time frames. Nevertheless, the chosen window length remains close to regulatory recommendations. Specifically, Basel III guidelines suggest that financial institutions adopt a minimum of 365 daily observations when computing Value-at-Risk (VaR) metrics for internal risk models, placing the current setting within proximity of those requirements.

It is acknowledged, however, that the relatively small window size may limit the model's ability to fully capture long-memory effects and volatility clustering, particularly during extended periods of market stress. Literature on GARCH-type models often suggests that optimal rolling windows for stable out-of-sample performance lie in the range of 1,000 to 1,500 observations. Given sufficient computational resources, future analyses would benefit from testing larger windows to enhance model robustness and reliability.

DCC-GARCH(1,1) VaR rolling window forecast

I further evaluate the robustness of the DCC-GARCH(1,1) model by performing a rolling one-step ahead forecast of the 5% Value-at-Risk (VaR) on the equally weighted portfolio. A rolling window of 300 observations was employed, updating recursively over the test set. At each step, the model was re-estimated on the most recent window, and the one-step ahead portfolio VaR was computed using the dynamically forecasted covariance matrix.

Figure 13 illustrates the predicted VaR alongside the actual portfolio returns and the observed violations. The out-of-sample violation rate is **7.03%**, which remains broadly consistent with the in-sample violation level (previously estimated around 4.85%). While the forecasted VaR slightly underestimates risk during a few volatile episodes, the overall performance confirms the model's reliability in capturing short-term downside risk in a multivariate setting. The relatively close alignment between in-sample and out-of-sample violation rates suggests that the DCC-GARCH model generalizes well and provides stable risk forecasts across time.

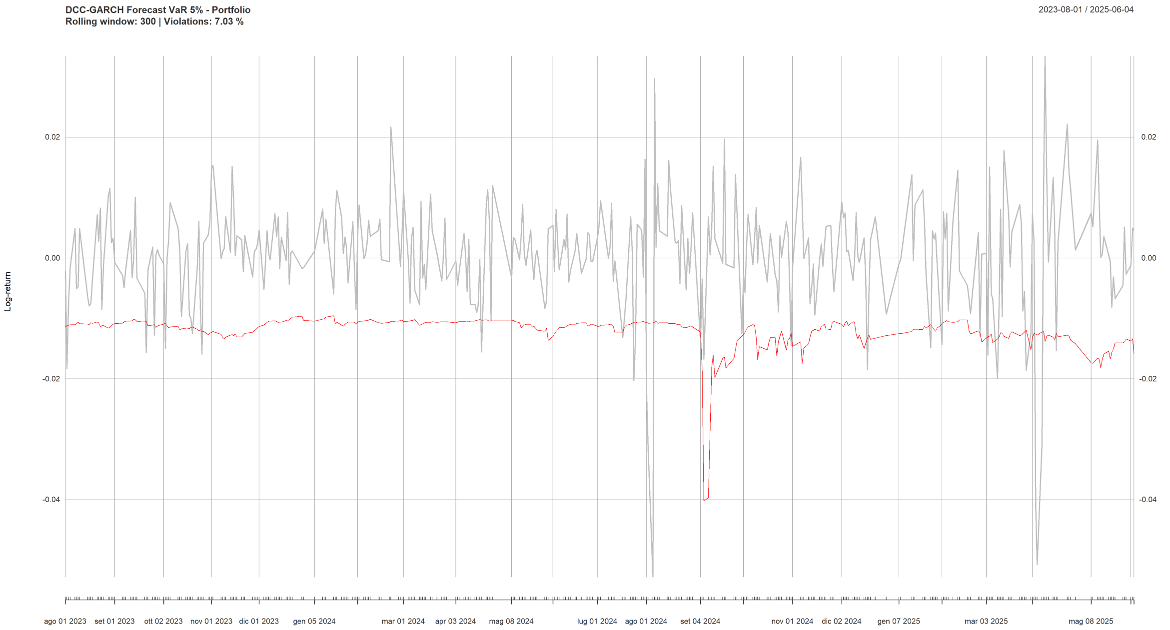


Figure 13: Rolling one-step ahead forecast of 5% VaR using DCC-GARCH(1,1) on an equally weighted portfolio. The red line shows the forecasted VaR and gray lines indicate actual log-returns

GARCH(1,1) VaR rolling window forecast

The out-of-sample evaluation of the GARCH(1,1) model on the equally weighted portfolio reveals a substantial discrepancy between expected and realized risk. As shown in Figure 14, the rolling one-step ahead 5% Value-at-Risk (VaR) forecast using GARCH(1,1) results in a violation rate of **12.24%**, significantly exceeding the theoretical benchmark of 5%. This underperformance aligns with the pattern observed during the in-sample phase, where the model systematically underestimated downside risk.

The primary reason for this result lies in the univariate nature of the GARCH(1,1) framework, which treats each asset in isolation and fails to account for time-varying correlations between portfolio components. As a consequence, the aggregated risk of the portfolio is underestimated—particularly during periods of market stress when asset returns tend to become more correlated. This structural limitation leads to frequent VaR violations and diminished reliability in real-world portfolio applications. The results thus reinforce the earlier conclusion that while GARCH(1,1) may capture individual volatility dynamics reasonably well, it is ill-suited for multivariate or portfolio-level risk forecasting tasks without an explicit modeling of interdependencies.

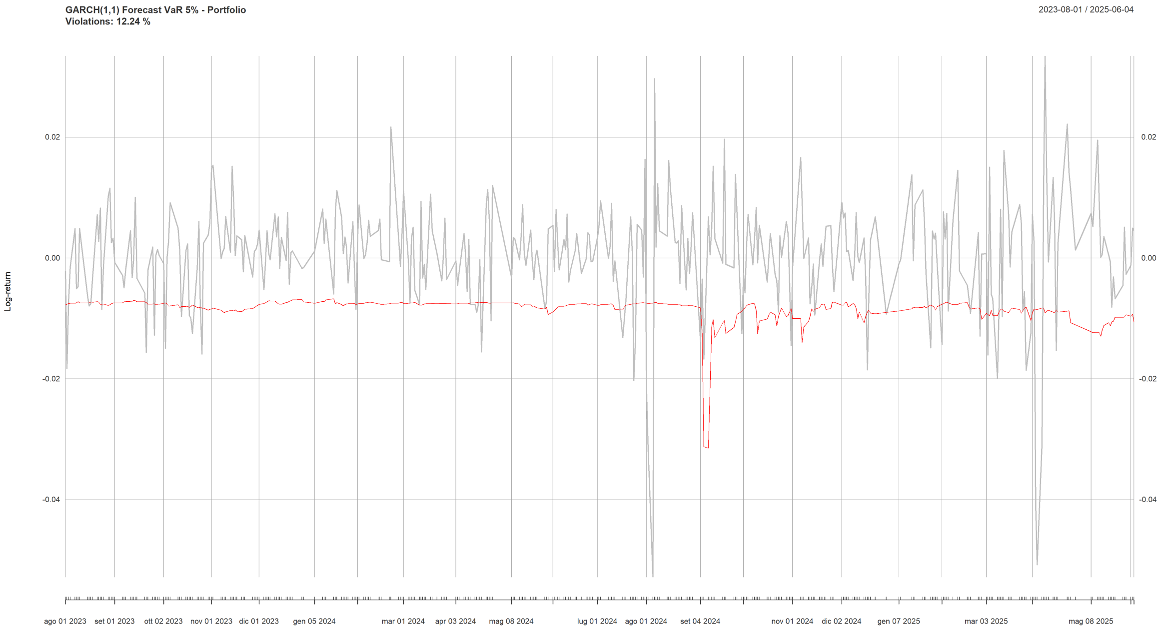


Figure 14: Rolling one-step ahead forecast of 5% VaR for the equally weighted portfolio using univariate GARCH(1,1). The red line indicates the VaR threshold. The violation rate is 12.24%, well above the expected 5%.

Conclusion

The empirical results presented in this study highlight a clear distinction in the performance of volatility models depending on the context of application. While univariate GARCH(1,1) models remain among the most widely used and effective tools for modeling and forecasting volatility in individual financial time series, their application to portfolio-level risk measurement proves to be considerably less reliable. The in-sample and out-of-sample Value-at-Risk (VaR) analysis demonstrates that the GARCH(1,1) model often underestimates downside risk when applied to an equally weighted portfolio, yielding an excessive number of VaR violations.

This discrepancy is fundamentally due to the model's inability to account for cross-asset correlations, particularly their time-varying nature, which becomes especially relevant during periods of heightened market volatility. In contrast, the DCC-GARCH(1,1) model—through its dynamic conditional correlation structure—captures both the individual conditional variances and the evolving co-movement among assets. As a result, portfolio VaR forecasts generated by the DCC-GARCH model exhibit violation rates much closer to the theoretical confidence level, thus offering a more accurate and robust representation of portfolio risk.

From a theoretical standpoint, the advantages of multivariate GARCH frameworks stem from their foundation in modern portfolio theory, which emphasizes the importance of variance-covariance structures in measuring total portfolio risk. The DCC-GARCH model, in particular, offers a practical balance between flexibility and parsimony, allowing for time-varying correlations without the computational complexity of full VEC or BEKK specifications.

In conclusion, while GARCH(1,1) remains a benchmark model for single-asset volatility analysis due to its simplicity and empirical tractability, multivariate extensions such as DCC-GARCH should be preferred for portfolio-level risk management. Their superior ability to capture joint dynamics and adapt to changing market conditions makes them more appropriate for real-world applications, especially in contexts characterized by systemic risk and cross-market contagion.

Technological choices and code

The project was totally done in Visual Studio Code (VSC); given the modest computational complexity and size of the code, I did not decide to implement code engineering policies via CI-CD or Agile development methodologies.
The code is made in single classes, developed in a single "ready to execute" script with no external classes written in S3, S4 and R6. The code therefore is a script and does not follow OOP rules.

Deployment of the code was done via GitHub (the major Version Control System) linked to Visual Studio, with a push to the main branch for each significative update of the code. A version of the code is available at the public repository on GitHub shown below.

> **Warning!**
> To download the file, go to the top right and press "code" and then "Download Zip "which also includes the dataset. Also the current working directory should be changed in the code.
>
> **Download from GitHub**