Proceedings of Business and Economic Studies, 2025, Volume 8, Issue 6
http://ojs.bbwpublisher.com/index.php/PBES
Online ISSN: 2209-265X
Print ISSN: 2209-2641

Quantitative Risk Modeling and Portfolio Construction with ARMA-GARCH: An Empirical Study on the S&P 500

Xiaoning Zhang*

University of Birmingham, Birmingham B5 4SJ, United Kingdom

**Author to whom correspondence should be addressed.*

Abstract: This study investigates the return dynamics, volatility structure, and risk characteristics of five representative S&P 500 stocks: Johnson & Johnson, Microsoft, NVIDIA, Coca-Cola, and Home Depot, using ARMA-GARCH models. Descriptive statistics and diagnostic tests confirm non-normality, negative skewness, fat tails, and volatility clustering, providing strong justification for conditional mean-variance modelling. Optimal model specifications are selected via the Bayesian Information Criterion, with EGARCH frameworks generally outperforming alternative GARCH variants in capturing asymmetric volatility responses. Rolling-window forecasts for 2024Q1 show that the models generate stable and reliable volatility predictions for low-volatility stocks (JNJ, KO), while performance is weaker for highly volatile stocks (NVDA), highlighting structural limitations under extreme market shifts. To evaluate risk management implications, one percent Value-at-Risk and expected shortfall were computed and backtested. Results indicated conservative tail-risk forecasts, with violation rates well within acceptable thresholds. Portfolio applications are further explored by constructing the Global Minimum Variance Portfolio (GMVP) and the Maximum Sharpe Ratio (Max SR) portfolio using rolling covariance estimates. Out-of-sample backtesting demonstrated that the GMVP delivered low volatility but modest returns, whereas the Max SR portfolio achieved significantly higher performance, consistent with the risk-return trade-off. Overall, the findings confirm that ARMA-GARCH models are effective tools for modelling conditional volatility and informing dynamic asset allocation. However, their limited adaptability to jump risk and nonlinear structural breaks underscores the need for more advanced modelling approaches in high-volatility environments.

Keywords: ARMA-GARCH; Expected shortfall; Portfolio optimization; S&P 500; Value-at-Risk; Volatility forecasting

Online publication: November 6, 2025

1. Data collection and description

1.1. Selection of stocks

As shown in **Table 1**, the five selected stocks constitute a representative portfolio encompassing growth, defensive, and cyclical characteristics, consistent with the evidence of common risk factors highlighted by Fama and French [1].

Table 1. Selection of stocks

Stock name	Ticker	Sector	Selection rationale
Johnson & Johnson	JNJ	Health care	Low-volatility, defensive stock with stable cash flow and strong profitability, making it well-suited for portfolio inclusion
Microsoft Co.	MSFT	Information technology	A leading technology blue-chip with robust fundamentals and consistent profitability, serves as a representative long-term growth stock
NVIDIA Co.	NVDA	Artificial intelligence	A dominant player in the semiconductor industry, is characterized by high volatility and significant market influence
Coca-Cola Co.	KO	Consumer staples	A classic defensive consumer stock, offers stable earnings and plays a key role in conservative asset allocation strategies
Home Depot Inc.	HD	Consumer discretionary	A high-growth, high-profitability retail stock, is an essential candidate for portfolio construction in consumer-driven sectors

1.2. Descriptive statistics of log returns

Descriptive statistics further support this characterization. As shown in **Table 2**, NVDA has the highest mean return (0.00105) and volatility (SD = 0.03746), confirming its high-risk/high-return profile. JNJ and KO offer lower returns and volatility, reinforcing their defensive nature. All stocks exhibit negative skewness, suggesting susceptibility to extreme negative returns. High kurtosis values indicate fat-tailed distributions, highlighting elevated tail risk.

Table 2. Statistics test of log returns

Stock	Mean	SD	Skewness	Kurtosis
JNJ	0.00030	0.01216	-0.475	17.305
MSFT	0.00039	0.01925	-0.151	12.215
NVDA	0.00105	0.03746	-0.211	16.191
KO	0.00023	0.01307	-0.199	12.264
HD	0.00036	0.01931	-1.092	26.396

Price trends illustrate that HD and MSFT achieved the strongest long-term growth. NVDA, despite a lower starting price, surged post-2016, reflecting its high-growth trajectory. JNJ and KO maintained stable prices with low volatility, underscoring their role as hedging assets (**Figure 1**).

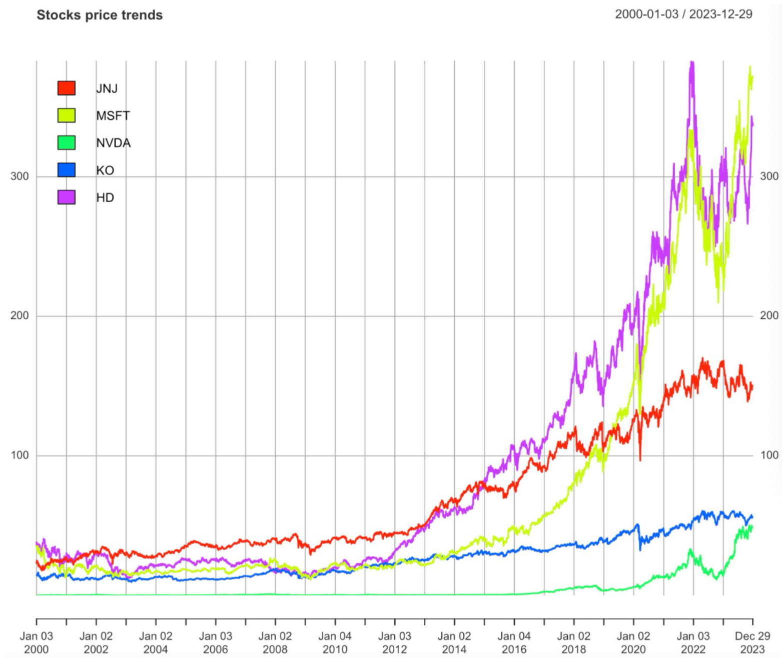

Figure 1. The price trends of varying stocks.

2. Properties of daily log returns

Statistical tests on the daily log returns of the five selected stocks reveal that mean returns are only weakly significant in some cases (e.g., NVDA: $P = 0.0300$; JNJ: $P = 0.0515$), suggesting limited predictive power based on mean alone (**Table 3**). All stocks showed significant skewness and kurtosis ($P = 0$), rejecting symmetry and normal tail assumptions, a result in line with the stylized facts of financial return distributions documented by Fama and French [1]. This implies leptokurtic distributions, which is typical in financial return data.

The Jarque-Bera test confirms these findings, implying non-normal return distributions with fat tails and asymmetry. The Ljung-Box test identifies strong serial correlation in raw returns, validating the use of ARMA models for the conditional mean. Moreover, all stocks show significant autocorrelation in squared returns ($P = 0$), indicating volatility clustering and justifying the use of GARCH-type models.

Table 3. Statistical tests for log returns

Stock	Mean (*P*)	Skew (*P*)	Kurt (*P*)	JB (*P*)	LBQ (*P*)	LBQ (sq-*P*)
JNJ	0.0515	0	0	0	0.0000	0
MSFT	0.1173	0	0	0	0.0000	0
NVDA	0.0300	0	0	0	0.0001	0
KO	0.1662	0	0	0	0.0088	0
HD	0.1492	0	0	0	0.0000	0

In summary, the empirical tests provide robust evidence that:

(1) The daily returns of all five stocks are non-normal, asymmetric, and fat-tailed;

(2) There is strong evidence of both serial correlation and volatility clustering;
(3) ARMA-GARCH models with appropriate distributional assumptions are well-suited for modeling the return dynamics of these stocks.

3. Modelling and forecasting variance and returns

3.1. Five firms optimal model selection

In the following part, ARMA-GARCH models were systematically constructed based on the daily log return series of the five selected companies. For each company, the optimal model was selected using the BIC to capture the dynamics of return and volatility.

During the mean equation modeling phase, I fit ARMA (p,q) models within the lag range (p,q) ∈ [0,2] to each company's return series:

$$r_t = \mu + \sum_{i=1}^{p} \phi_i r_t - i + \sum_{j=1}^{q} \theta_j \varepsilon_t - j + \varepsilon_t$$

Here, ϕ_i denotes the AR coefficients, θ_j the MA coefficients, and ε_t is the white noise error term.

$$BIC = -2\ln\left(\hat{L}\right) + k\ln\ (n)$$

BIC was used as the model selection criterion to ensure a balance between model fit and complexity. Residual diagnostics are then conducted to examine the presence of heteroskedasticity.

As shown in **Table 4**, residuals generally exhibit negative skewness and leptokurtosis. The Jarque-Bera test rejects normality at the 1% significance level, while the Ljung-Box test indicates significant autocorrelation (*P*-values well below 0.05), justifying the use of GARCH-type models.

Table 4. Residuals diagnostics for ARMA models

Stock	Skewness	Kurtosis	JB-Test	LB-Test
JNJ	-0.5381	17.1535	0	8e-04
MSFT	-0.1695	12.0324	0	7e-04
NVDA	-0.2111	16.1913	0	1e-04
KO	-0.1989	12.2637	0	0.0088
HD	-1.0916	26.3962	0	0

In the volatility modeling phase, I consider SGARCH, GJR-GARCH, and EGARCH frameworks, estimated using the rugarch package in R [2].Using BIC for model selection under the assumption of normal distribution. The optimal GARCH (p,q) structure is searched within (p,q) ∈ [1,2].

For example, in a first-order model:

(1) SGARCH (1,1):

$$r_t = \mu + \varepsilon_t, \varepsilon_t = \sigma_t z_t$$
$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

(2) GJR-GARCH (1,1):

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 I\left(\varepsilon_{t-1} < 0\right) + \beta\sigma_{t-1}^2$$

(3) EGARCH (1,1):

$$\log\left(\sigma_t^2\right) = \omega + \beta\log\left(\sigma_{t-1}^2\right) + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \alpha\left(\left|\frac{\varepsilon_{t-1}}{\varepsilon_{t-1}}\right| - \sqrt{\frac{2}{\pi}}\right)$$

σ_t^2 represents conditional variance, α measures the immediate impact of return shocks on volatility, and β captures volatility persistence.

The final model selection results for the five companies are summarized in **Table 5**.

Table 5. Optimal ARMA-GARCH models for each stock

Stock	Best ARMA model (p,q)	BIC (ARMA)	Best conditional variance model	BIC (GARCH)
JNJ	ARMA(0,2)	-36097.63	EGARCH(1,1)	-6.3097
MSFT	ARMA(0,1)	-30569.52	EGARCH(2,1)	-5.3746
NVDA	ARMA(0,0)	-22504.64	EGARCH(1,2)	-4.0486
KO	ARMA(0,0)	-35217.35	EGARCH(1,1)	-6.1968
HD	ARMA(0,0)	-30506.04	EGARCH(1,1)	-5.4885

Overall, most stocks exhibit white noise characteristics in their return processes (typically ARMA (0,0) or similar), while their volatility processes show clear asymmetries and lagged shock effects. Among the models considered, EGARCH generally offers superior fitting performance compared to SGARCH and GJR-GARCH.

3.1.1. JNJ ARMA (0,2)-EGARCH (1,1)

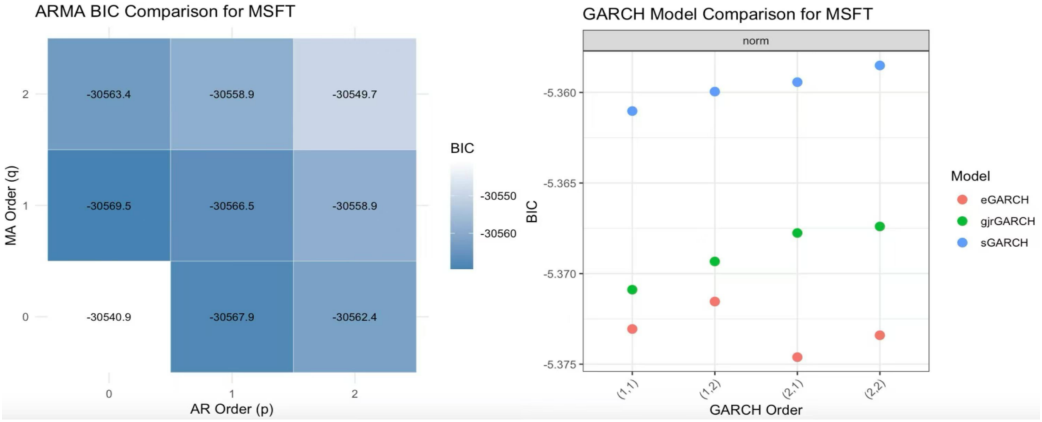

Figure 2. ARMA-GARCH model comparison for JNJ.

3.1.2. MSFT ARMA (0,1)-EGARCH (2,1)

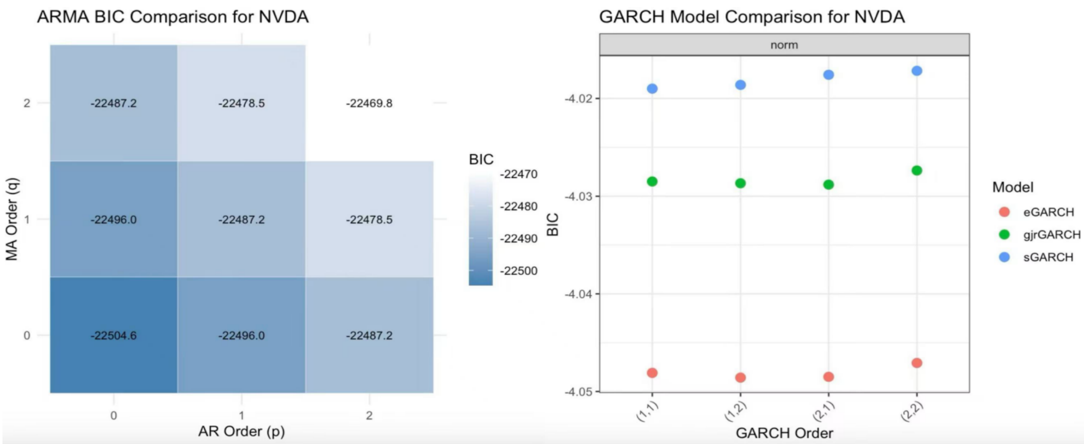

Figure 3. ARMA-GARCH model comparison for MSFT.

3.1.3. NVDA ARMA (0,0)-EGARCH (1,2)

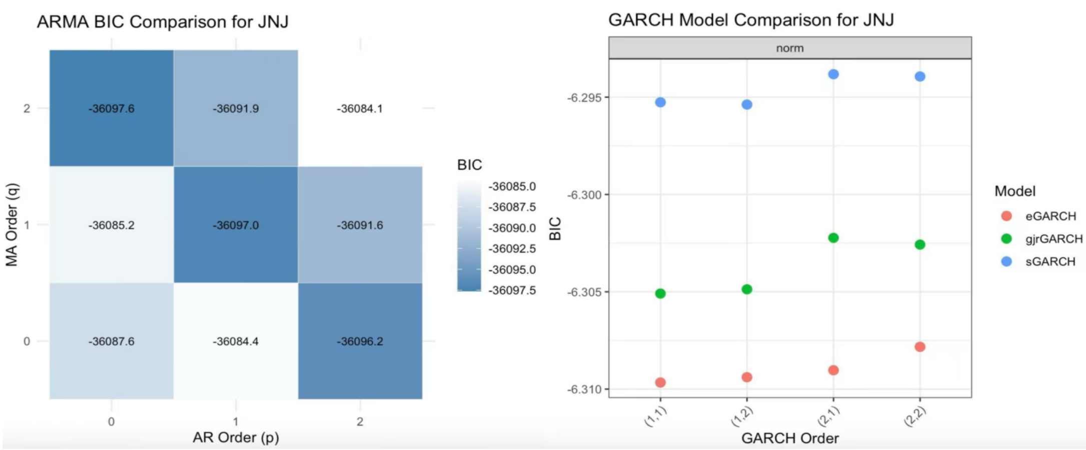

Figure 4. ARMA-GARCH model comparison for NVDA.

3.1.4. KO ARMA (0,0)-EGARCH (1,1)

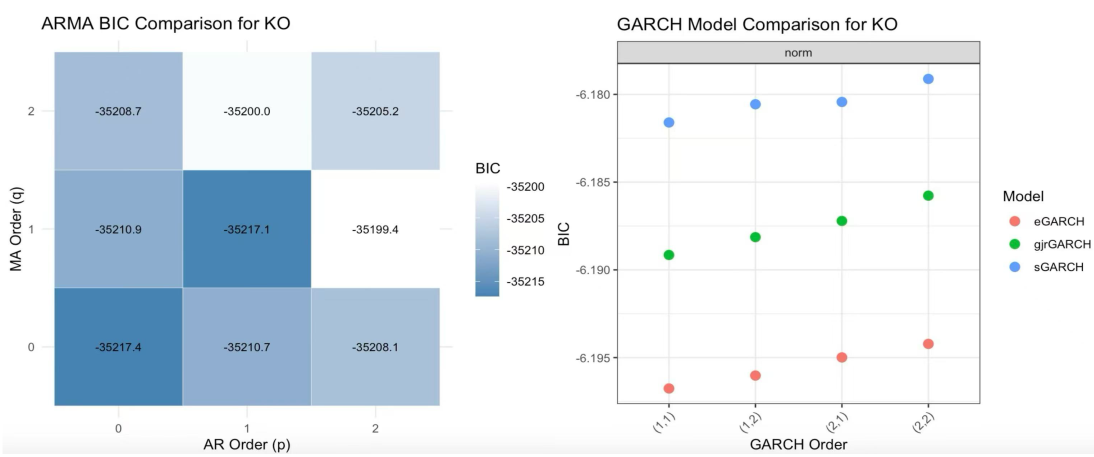

Figure 5. ARMA-GARCH model comparison for KO.

3.1.5. HD ARMA (0,0)-EGARCH (1,1)

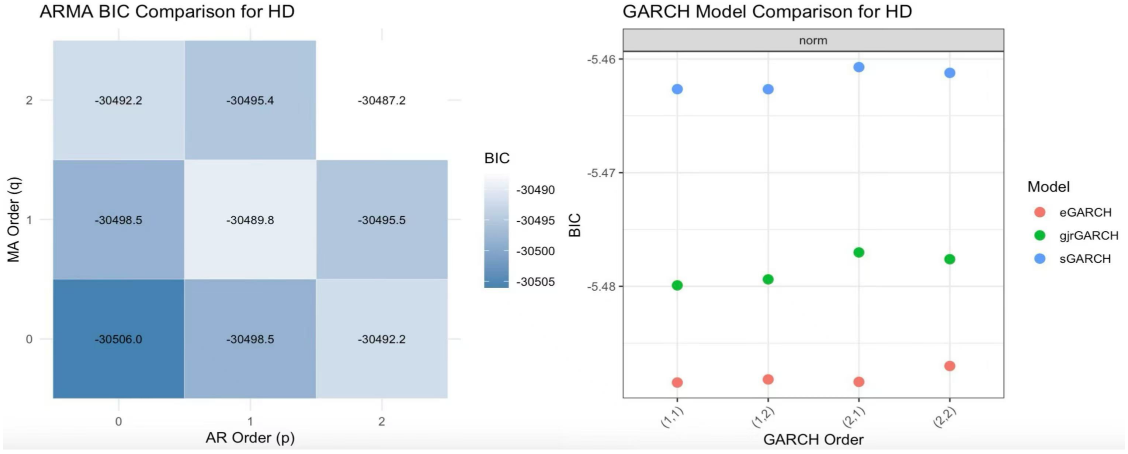

Figure 6. ARMA-GARCH model comparison for HD.

3.2. Forecast daily variance and return for each stock

A rolling window forecasting method was applied to model and predict the daily returns and volatilities of five stocks. To systematically assess the predictive accuracy of the ARMA-GARCH models, the predicted and actual values of return and volatility for 2024 Q1were compared. The key observations are as follows:

3.2.1. JNJ

As shown in **Figure 7**, the forecasted volatility remains persistently high but generally aligns with actual trends, suggesting the model produces stable daily forecasts but cannot quickly adapt to sudden market shocks.

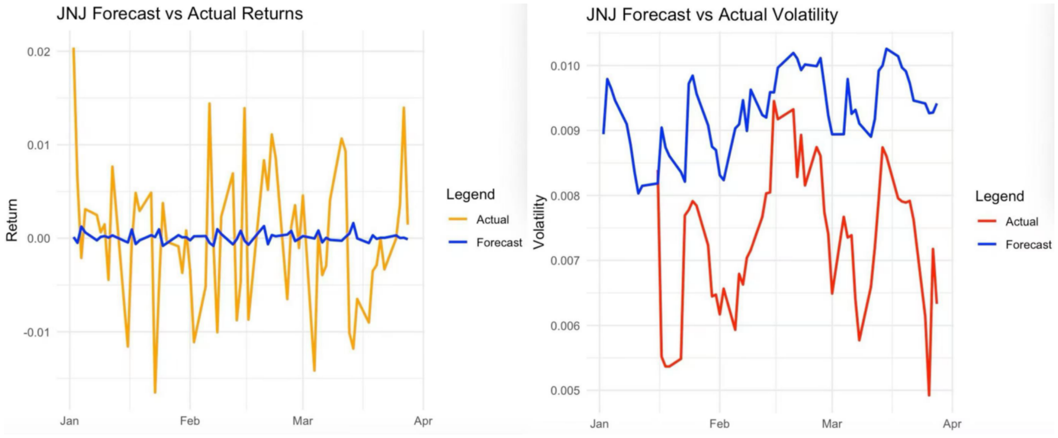

Figure 7. JNJ forecast vs actual returns (left) & volatility (right).

3.2.2. MSFT

Figure 8 indicates that volatility forecasts track the actual values relatively well.

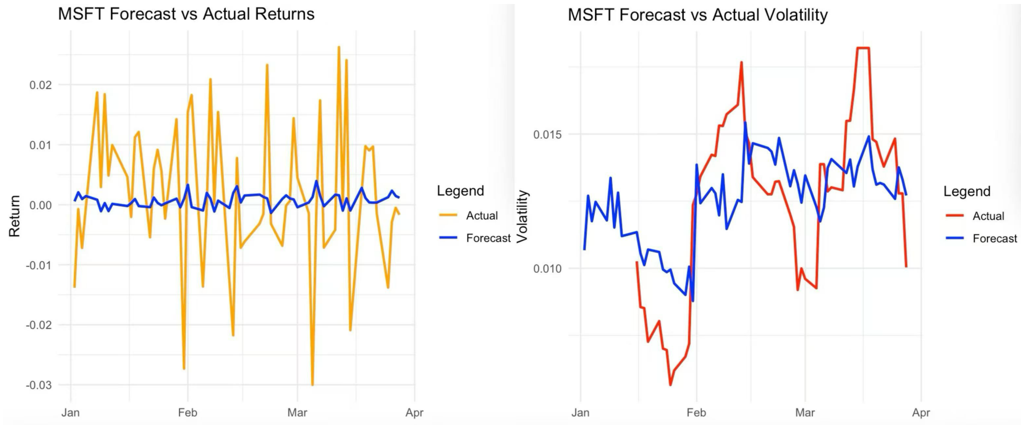

Figure 8. MSFT forecast vs actual returns (left) & volatility (right).

3.2.3. NVDA

As seen in **Figure 9**, actual returns are highly volatile, while forecasted returns show a monotonic upward trend. This indicates that the model struggles with capturing daily fluctuations.

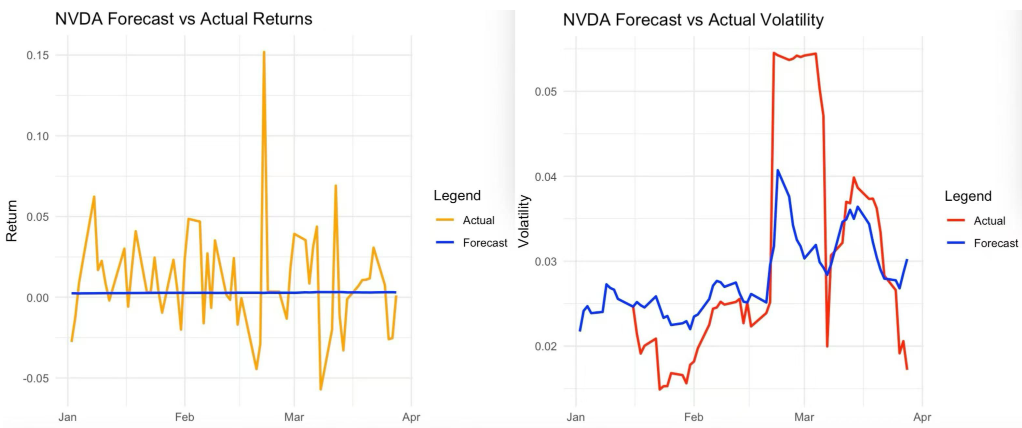

Figure 9. NVDA forecast vs actual returns (left) & volatility (right).

3.2.4. KO

As shown in **Figure 10**, the volatility forecast generally matches actual trends. However, peak timing mismatches suggest lagging adjustment to volatility shifts.

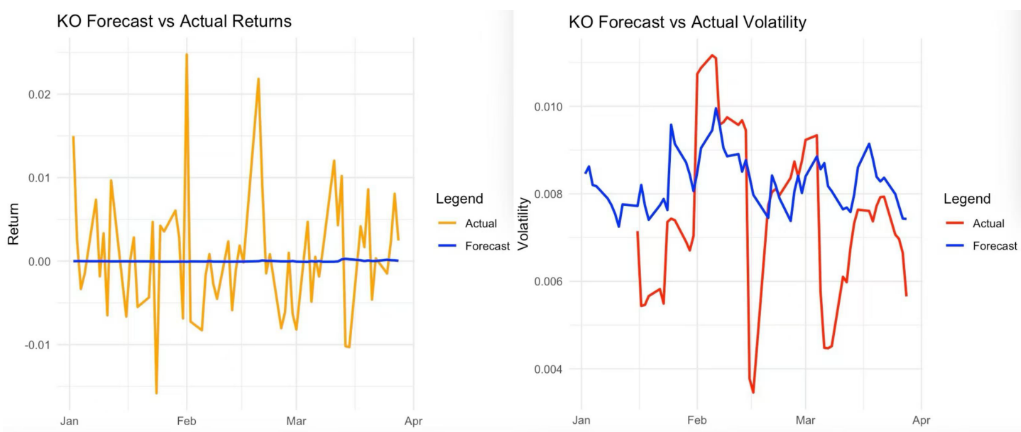

Figure 10. KO forecast vs actual returns (left) & volatility (right).

3.2.5. HD

Volatility forecasts match well in earlier periods, providing some early warning capacity. However, mismatches in later stages suggest lag in capturing real-time shifts (**Figure 11**).

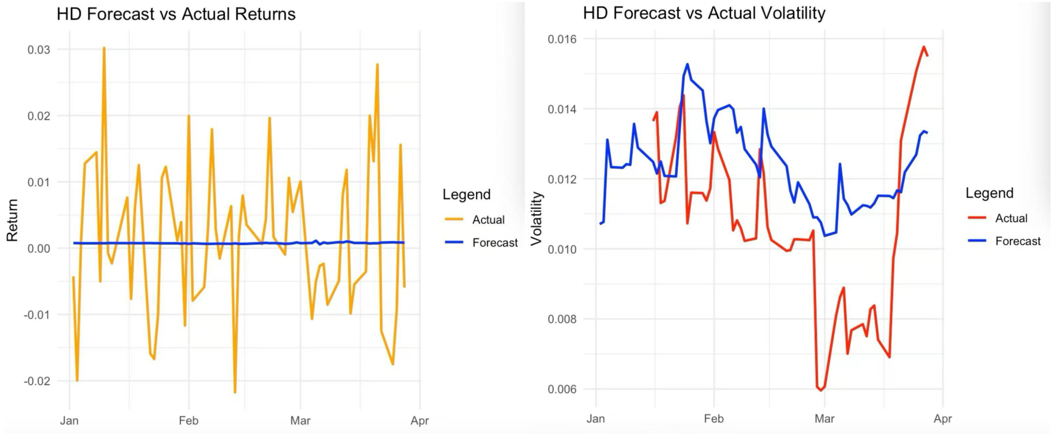

Figure 11. HD forecast vs actual returns (left) & volatility (right).

In summary, the selected ARMA-GARCH models demonstrate a certain degree of capability in identifying and fitting the daily volatility structure of the five stocks, especially in providing stable volatility forecasts. However, their ability to adapt to structural changes in actual returns is limited, indicating a need for more sophisticated models.

The performance evaluation metrics (RMSE and MAE) further confirm these findings:

(1) Models for JNJ and KO are stable and well-fitted;

(2) MSFT and HD perform reasonably with acceptable error levels;

(3) NVDA, due to its high-growth and jump-prone characteristics, requires a more advanced modeling approach.

Table 6. Prediction performance evaluation

Stock	JNJ	MSFT	NVDA	KO	HD
RSME	0.007559139	0.012033189	0.031644395	0.007295385	0.011377432
MAE	0.005939529	0..009358054	0.021840805	0.005446621	0.009195828

4. Value-at-Risk (VaR) calculation

The risk measurement performance of the ARMA-GARCH models was assessed by computing and backtesting the 1% quantile VaR and ES for each stock. For each stock, the 1% VaR was calculated under the assumption of normality using the rolling forecasts of conditional mean and conditional volatility. The formula is:

$$V\alpha R_1\% = \mu_t + \sigma_t \cdot z_\alpha$$

where μ_t is the forecasted return, σ_t is the forecasted volatility, and z_α is the 1% quantile of the standard normal distribution (approximately 2.326)

The following presents the VaR backtesting performance of five stocks during 2024 Q1:

Overall, the actual returns of all stocks, except MSFT, predominantly stayed above the VaR thresholds, indicating that the models were generally conservative and breaches were rare. Volatility levels varied significantly across sectors, with NVDA being the most volatile and KO the most stable.

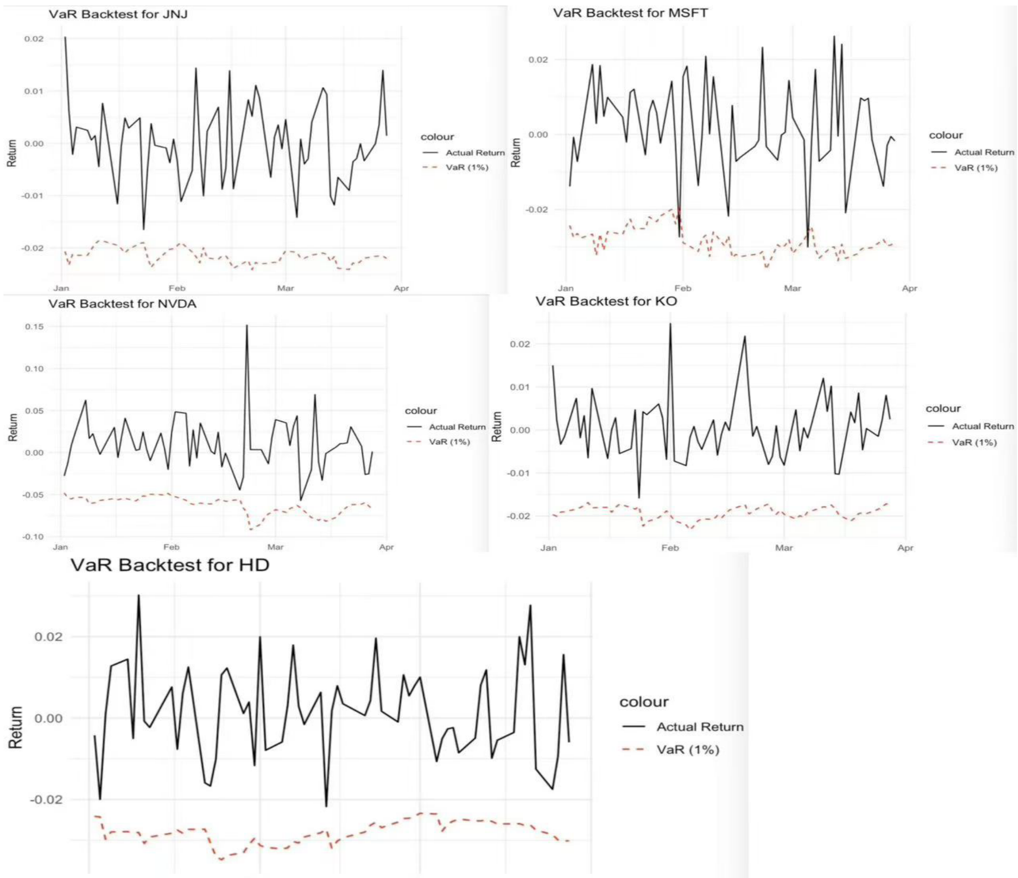

Figure 12. Five stocks 1% value-at-risk.

Following this, the corresponding ES is calculated as:

$$ES_1\% = \mu_t - \frac{\sigma_t \cdot \phi\ (z_\alpha)}{\alpha}$$

where ϕ (z_α) is the standard normal density evaluated at z_α $(\alpha = 0.01)$.

This provides a more accurate measure of tail risk.

To evaluate model accuracy and stability, the POF test is applied following Kupiec [3]. The results are as shown in **Table 7**, indicating that most models exhibit low violation rates and do not reject the null hypothesis, suggesting the ARMA-GARCH models effectively capture tail risk at the 1% level.

Table 7. VaR backtesting summary

Stock	Violation Rate	Kupiec (*P*)	ES (sum)
JNJ	0.0000	0.2682	-1.504830
MSFT	0.0328	0.1571	-1.986698
NVDA	0.0000	0.2682	-4.389194
KO	0.0000	0.2682	-1.345106
HD	0.0000	0.2682	-1.971365

Each stock's optimal ARMA-GARCH specification was used to perform rolling forecasts, yielding the following average 1% VaR estimates as shown in **Table 8**.

Table 8. 1% VaR estimate results

Stock	1% VaR estimate
JNJ	-2.15%
MSFT	-2.83%
NVDA	-6.24%
KO	-1.92%
HD	-2.81%

Based on the results, NVDA has a significantly higher 1% VaR than other stocks, indicating potential daily losses of 6.24% or more under extreme market conditions, reflecting higher volatility and greater tail risk. KO and JNJ show the lowest VaR, consistent with the stable nature of defensive stocks. MSFT and HD exhibit similar VaR levels, suggesting comparable risk profiles during the modeling period despite differing industries. These findings inform risk management, where high-VaR stocks may require tighter position control or hedging strategies, while low-VaR stocks can serve as risk buffers in a diversified portfolio.

5. Portfolio analysis

Subsequently, a rolling window estimation and out-of-sample backtesting method was applied to construct and evaluate two classic portfolio strategies (GMVP and Max SR) under the real market environment of 2024 Q1.

GMVP aims to minimize the overall portfolio volatility, the weight vector for the GMVP is calculated as:

$$\omega GMVP = \frac{\Sigma^{-1} 1}{1^{'} \Sigma^{-1} 1}$$

where Σ^{-1} is the inverse of the covariance matrix and **1** is a column vector of ones.

While the Max SR portfolio seeks to maximize the excess return per unit of risk given a risk-free rate, its optimal weights can be represented as:

$$\omega_{MaxSR} = arg \max_{w} \frac{\omega^{'} (\mu - r_f)}{\sqrt{\omega^{'} \Sigma w}}$$

where r_f denotes the risk-free rate.

Taking into account practical market conditions and optimal portfolio requirements, the following constraints were introduced for both strategies:

(1) Non-negative weights(no short selling) with the total weight not exceeding 100%;

(2) The risk-free rate is set at 2%, based on the U.S. Treasury yield hovering around this level in late 2023 and early 2024, ensuring both realism and the stability of the Sharpe ratio calculation;

(3) A maximum allocation of 30% per single stock, to prevent over-concentration and reduce the impact of individual stock volatility on the overall portfolio.

For each out-of-sample date, to ensure stable and reliable estimation of expected returns (μ_t) and covariance matrices (Σt), 1000 trading days were rolled back, as this window length balances the trade-off between the need for sufficient historical information and the relevance to current market conditions, using maximum likelihood estimation:

$$\widehat{\mu_t} = \frac{1}{M}\sum_{i=1}^{M} r_{t-i}$$

$$\widehat{\Sigma_t} = \frac{1}{M}\sum_{i=1}^{M} (r_{t-i} - \widehat{\mu_t})(r_{t-i} - \widehat{\mu_t})'$$

where M = 1000, and r_{t-i} is the log return on day *t-i*.

Based on the estimation results, the portfolio weights were re-optimized, and rebalanced daily to form a complete sequence of out-of-sample portfolio returns. The backtest results are shown in **Table 9**.

Table 9. Assets allocation weight table

Stock	GMVP	Max SR
JNJ	30%	0%
MSFT	16.2%	30%
NVDA	0%	30%
KO	30%	20%
HD	23.8%	30%

As shown in **Table 10**, the GMVP portfolio concentrated its holdings in JNJ, KO, and HD, while completely avoiding NVDA. This allocation is consistent with the GMVP's objective of minimizing volatility, given that NVDA exhibited extremely high volatility, as identified in the VaR analysis.

In contrast, the Max SR portfolio, aiming to maximize the Sharpe ratio, preferred high-return assets despite higher volatility. MSFT, NVDA, and HD were each assigned the maximum allowed weight of 30%, with NVDA heavily overweighted to capitalize on its high expected returns.

Table 10. Return performance of the two portfolios

Portfolio	Total-Return ($E[R_p]$)	Annualized-Vol (σ_p)	Sharp ratio
GMVP	6.34%	9.3%	0.43
Max SR	27.86%	20.97%	0.82

The results further highlight the differences in performance: the GMVP achieved an annualized volatility of 9.3% and a relatively low Sharpe ratio of 0.43, with a total return of 6.34%, demonstrating its low-risk, low-return nature. In contrast, the Max SR portfolio exhibited a higher annualized volatility of 20.97%, but achieved a substantially higher total return of 27.86%.

Based on these figures, the theoretical annualized Sharpe ratio without adjustment would be:

$$\text{Sharpe Ratio} = \frac{E[R_p] - R_f}{\sigma_p} = \frac{27.86\% - 2\%}{20.97\%} \approx 1.23$$

However, based on the R code, the Sharpe ratio is adjusted for the actual number of trading days in the out-of-sample period using the following scaling factor:

$$\text{Adjusted Sharpe Ratio} = \left(\frac{\overline{R_p} - \frac{R_f}{252}}{\sigma_p}\right) \times \sqrt{\frac{252}{(Period\ Days/21) \times 12}}$$

This adjustment corrects for the short evaluation window(three months) to avoid overstating performance. As a result, the final Sharpe ratio for the Max SR portfolio is 0.82. Although the Max SR portfolio assumed greater volatility, it significantly outperformed the GMVP on a risk-adjusted basis.

As illustrated in **Figure 13**, the Max SR portfolio displayed a consistently upward trajectory, with cumulative returns rising sharply and exceeding 30% within three months. Meanwhile, the GMVP portfolio maintained a much steadier and flatter trajectory, with cumulative returns around 5%.

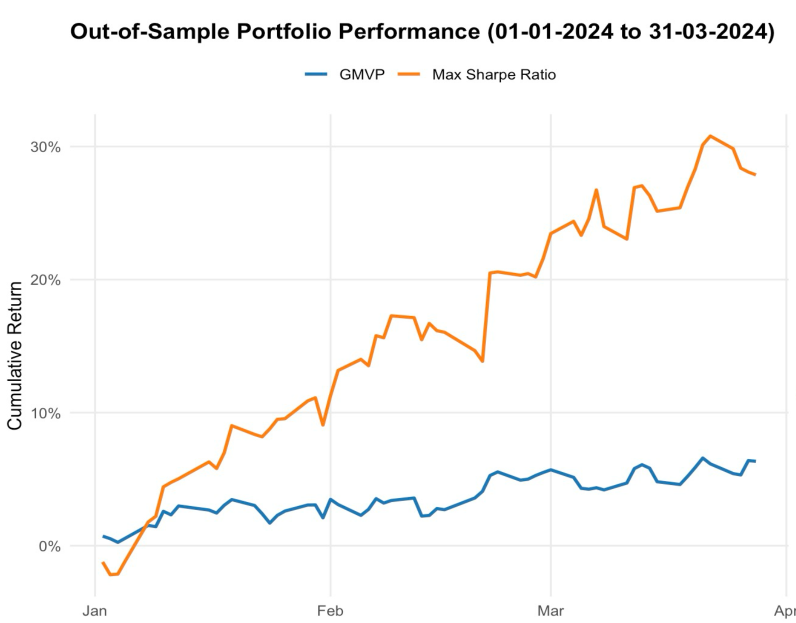

Figure 13. Out-of-sample portfolio performance 2024 (Q1).

From an economic perspective, during Q1 2024, the Federal Reserve signaled a potential rate cut, and the AI and technology sectors experienced a strong rally. The Max SR portfolio, by overweighting high-expected-return assets such as MSFT and NVDA during optimization, captured the upward momentum in the market and significantly outperformed the GMVP [1].

Overall, the out-of-sample backtesting results validate the effectiveness of the two portfolio strategies. Each portfolio exhibited characteristics consistent with its intended risk-return trade-off, demonstrating robustness and economic plausibility under real-world market conditions.

6. Discussion and interpretation

6.1. The statistical properties of the selected stocks

In summary, the five selected stocks in this study constitute a representative portfolio encompassing growth, defensive, and cyclical characteristics. Based on empirical test results, all return series exhibit non-normality, negative skewness, and fat tails, indicating significant tail risks.

NVDA and MSFT show high market sensitivity and aggressive investment styles, while JNJ and KO demonstrate strong profitability and conservative investment behavior, making them suitable as risk-buffering assets in a portfolio. Moreover, the Ljung-Box test reveals significant autocorrelation and ARCH effects in both return and squared return series for all five stocks, providing a sound justification for the application of ARMA-GARCH models.

6.2. Effectiveness of ARMA-GARCH models

ARMA-GARCH models were selected for each stock based on the BIC, demonstrating strong performance in capturing return autocorrelation and conditional heteroskedasticity. In mean process modeling, most stocks approximate white noise, with low-order ARMA models being sufficient. In volatility modeling, EGARCH models significantly outperformed SGARCH and GJR-GARCH in capturing asymmetry and lagged shock effects.

The selected models fit well in-sample, with standardized residuals showing no significant autocorrelation or heteroskedasticity. Out-of-sample results confirm that the models provide stable volatility forecasts for low-volatility assets like KO and JNJ. However, the models are less effective at handling sudden jumps and nonlinear structural changes in high-volatility stocks like NVDA, suggesting a need for more complex or jump-diffusion models. Rolling forecast results show that predicted volatility generally aligns with actual trends. The 1%VaR estimates closely match actual losses, with the models generally producing conservative forecasts. While this enhances reliability in risk estimation, it may come at the cost of reduced capital efficiency and return potential.

6.3. Portfolio performance and management implications

Using the variance-covariance matrices derived from the ARMA-GARCH rolling forecasts, two classic portfolio strategies were constructed: the GMVP and the Max SR. Out-of-sample backtesting revealed distinct yet theoretically consistent performances. The GMVP exhibited low annualized volatility (9.3%) and a total return of 6.34%, reflecting a low-risk, low-return profile. In contrast, the Max-Sharpe portfolio favored high-return assets (e.g., NVDA, MSFT), achieving a higher total return of 27.86%, higher volatility (20.97%), and a superior Sharpe ratio of 0.82, indicating stronger risk-adjusted performance.

7. Conclusion

These results validate the classic risk-return trade-off: Max SR is better suited to risk-seeking investors, while GMVP aligns with risk-averse preferences and confirms the ability of the ARMA-GARCH model to dynamically describe the volatility and the co-movement relationship between assets in dynamic asset allocation. In practice, different macroeconomic cycles may warrant different portfolio strategies based on individual risk tolerance.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Fama E, French K, 1993, Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics, 33(1): 3–56.

[2] Kupiec P, 1995, Techniques for Verifying the Accuracy of Risk Measurement Models. Journal of Derivatives, 3(2): 73–84.

[3] Ghalanos A, 2023, Rugarch: Univariate GARCH Models. R Package Version, 1.5–4.