Project Report
Master's degree in financial Modeling

Market Risk Analysis and Forecasting for JPMorgan Chase, AIG, and Goldman Sachs: A Comparative Study Using VaR and ES (HS and GARCH)

Subject: Financial Econometrics
Professor: Sullivan HUÉ
Students: Amidath ALEGUE | Simey HIN | Lydie Pudie ZOCLI
Academic Year: 2024 - 2025

Abstract

This project examines the market risk of three prominent financial institutions: JPMorgan Chase (a depository bank), American International Group (an insurance company), and Goldman Sachs (a broker-dealer) over the period 2007–2023. Using daily return data, we estimate and forecast both Value-at-Risk (VaR) and Expected Shortfall (ES) through two methodologies: Historical Simulation (HS) and variations of GARCH models, including the standard GARCH (1,1) as well as EGARCH (1,1), which accounts for asymmetric volatility. A rolling window approach is applied with two windows of 250 days and 500 days, enabling dynamic updates to risk measures and capturing shifts in market volatility.

Particular focus is placed on critical periods of financial distress, such as the 2008 Global Financial Crisis and the 2020 COVID-19 market crash. Backtesting procedures are implemented to assess the accuracy and reliability of the models, comparing forecasted VaR and ES levels against realized returns. This comprehensive analysis sheds light on the distinct risk profiles of the three institutions and evaluates the effectiveness of HS and advanced GARCH variations in identifying and quantifying market risk during periods of heightened volatility.

Keywords: GARCH (1,1), EGARCH (1,1), Historical Simulation (HS), Value-at-Risk (VaR), Expected Shortfall (ES), Volatility, Rolling Window, Forecasting, Backtesting, 2008 Global Financial Crisis, 2020 COVID-19.

Table of Contents

List of tables

List of figures

1. INTRODUCTION

The ability to quantify and manage market risk is a cornerstone of financial stability, as highlighted by authors like Jorion (2007) and Hull (2018), who emphasize the importance of robust risk models in periods of financial distress. The 2008 Global Financial Crisis and the 2020 COVID-19 market crash showcased the catastrophic impact of insufficient risk assessment tools, reinforcing the findings of Brownlees and Engle (2016) on dynamic market volatility. This project investigates market risk for three financial institutions, leveraging methodologies like Historical Simulation and GARCH models to provide a comprehensive understanding of their risk profiles.

2. LITERATURE REVIEW

The study of market risk measurement and management has been a central focus in financial research, particularly after major crises such as the Global Financial Crisis of 2008 and the COVID-19 pandemic. Authors like J.P. Morgan (1996) introduced the concept of Value-at-Risk (VaR) as a standard for measuring market risk. VaR, while widely used, has been critiqued for its inability to account for tail risks during extreme events, as highlighted by Artzner et al. (1999), who proposed Expected Shortfall (ES) as a coherent risk measure.

Historical Simulation (HS) and parametric models like GARCH are commonly employed for forecasting VaR and ES. HS is straightforward and relies on past data without making strong parametric assumptions (Pritsker, 2006), but it can fail to adapt to changing market conditions. In contrast, the GARCH (1,1) model introduced by Bollerslev (1986) effectively captures volatility clustering, a characteristic of financial time series, as noted by Engle (1982) in his ARCH framework. Studies by Alexander (2008) and McNeil et al. (2015) further emphasize the advantages of GARCH models in dynamic risk forecasting.

Recent works have explored the application of rolling windows for risk forecasting, allowing models to adapt to evolving market dynamics. Research by Kupiec (1995) and Christoffersen (1998) has established robust backtesting frameworks to evaluate the accuracy of risk models. They argue that both unconditional and conditional coverage tests are essential for assessing the reliability of VaR and ES forecasts.

By leveraging insights from these foundational studies, this project applies HS and GARCH (1,1) models to assess the market risk of three major financial institutions like JPMorgan Chase, American International Group, and Goldman Sachs over the period 2007–2023. This literature forms the theoretical underpinning for our investigation into the effectiveness of these models during periods of heightened volatility.

3. DATA AND METHODOLOGY

3.1. Data

The report uses daily adjusted closing prices for three large financial firms, including JPMorgan Chase (a depository bank), American International Group (an insurance company), and Goldman Sachs (a broker-dealer), collected from Yahoo Finance between 2007 and 2023. The dataset includes important events of financial crisis, such as the 2008 Global Financial Crisis and the COVID-19 pandemic of 2020. Log returns are calculated using these prices to reduce trends and standardize volatility over time. These returns provide the foundation for risk estimation and forecasting.

3.2. Methodology

To quantify and forecast market risk, Value-at-Risk (VaR) and Expected Shortfall (ES) were estimated using two approaches: Historical Simulation (HS) and GARCH-based models.

For HS method, we estimate risk levels using past return distributions without assuming a specific probability distribution. For GARCH based method, we compare different model for robustness and select the best-fitting model to capture conditional volatility dynamics, allowing for more responsive risk estimation. The best-fitting GARCH model was selected based on performance metrics such as Log-likelihood Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), ensuring consistency in risk forecasts across different model specifications. Furthermore, the diagnostic check, Ljung-Box, JB Test on standardized residual will be performed on the best-fitting model to check and validate the model. For each rolling window, these models were fit to the returns, and the conditional volatility forecasted for the subsequent day was used to compute VaR and ES.

GARCH Based Estimation

The GARCH (1,1) model used in this study is defined as:

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$$

Additionally, we will employ the extension model EGARCH (1,1) model if properties of the series suggested, which accounts for asymmetric volatility effects:

$$\ln(\sigma_t^2) = \omega + \alpha \ln(r_{t-1}^2) + \alpha \frac{r_{t-1}}{\sigma_{t-1}} + \gamma \frac{r_{t-1}}{\sigma_{t-1}}$$

where σ_t^2 represents the conditional variance at time, ω is a constant, α measures the impact of past squared returns, and β represents the persistence of past volatility, γ captures the leverage effect, reflecting the tendency of volatility to increase more after negative shocks than positive ones.

Rolling Window Framework: Two rolling window sizes were utilized for both HS and GARCH approaches:

- **250 trading days:** Representing approximately one year of trading data.
- **500 trading days:** Representing approximately two years of trading data.

This framework allowed for daily forecasts of VaR and ES, ensuring sensitivity to recent market movements and capturing changes in market volatility over time.

3.3. Backtesting

To evaluate the accuracy of the risk forecasts, back testing procedures were implemented:

- **Unconditional Coverage (UC) Test:** Assesses whether the proportion of observed exceedances matches the expected confidence level (e.g., 5% for a 95% confidence level).
- **Independence (IND) Test:** Evaluates whether exceedances are independent over time, ensuring no clustering of violations.
- **Conditional Coverage (CC) Test:** Combines the UC and IND tests to jointly assess the proportion and independence of exceedances.

By combining Historical Simulation and GARCH models (including EGARCH variants) with robust backtesting procedures, this methodology provides a detailed analysis of the risk profiles of JPMorgan Chase (a depository bank), American International Group (an insurance company), and Goldman Sachs (a broker-dealer). It evaluates the effectiveness of these methods in identifying and quantifying market risk, particularly during periods of heightened volatility, such as the 2008 Global Financial Crisis and the 2020 COVID-19 pandemic. This comprehensive approach highlights the adaptability of GARCH models and their variants in capturing evolving risk dynamics compared to the static nature of Historical Simulation.

4. TIME SERIES ANALYSIS

To analyze the dynamics of financial returns, it is essential to examine their statistical and temporal properties. This section focuses on analyzing time series by assessing their descriptive statistics, stationarity, and stylized facts related to variability, in order to better understand the underlying behaviors of the data.

4.1. Descriptive statistics

The table 1 presents the descriptive statistics of the returns for three companies: JPM (JPMorgan), AIG (American International Group), and GS (Goldman Sachs).

Table 1: Descriptive statistics

	Mean	Variance	Skewness	Kurtosis
JPM	0.0003	0.0006	0.2661	19.1751

AIG	-0.0007	0.0017	-3.0978	100.4868
GS	0.0002	0.0005	0.1875	17.8210

The mean returns indicate that JPM and GS have slightly positive returns, while AIG shows a negative average, suggesting a downward trend. The variance, which measures return volatility, is highest for AIG, indicating greater instability, whereas GS has the lowest variance, reflecting more stable returns. Skewness shows that the return distributions of JPM and GS are slightly right skewed, suggesting a higher probability of extreme gains, while AIG exhibits strong negative skewness, indicating a greater risk of extreme losses. Lastly, kurtosis, which measures the presence of extreme events, is particularly high for AIG, signaling a strong tendency for market shocks. These results suggest that AIG is the riskiest stock in the panel, with significant volatility and frequent extreme movements, whereas GS appears to be the most stable option for a risk-averse investor.

4.2. Stationarity Analysis of Stock Returns

To assess the stationarity of the stock returns, we conducted the **Augmented Dickey-Fuller (ADF) test** for three stocks: **JPM (JPMorgan), AIG (American International Group), and GS (Goldman Sachs)**.

Table 2: Augmented Diskey-Fuller (ADF) test

Stocks	**ADF_p-value**
JPM	0,01
AIG	0,01
GS	0,01

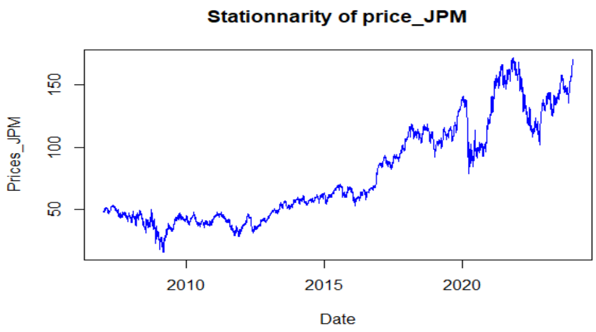
The results indicate that all three stocks have a **p-value of 0.01**, which is below the conventional significance levels (0.05). Given these results, we **reject the null hypothesis** of the ADF test, which states that the time series has a **unit root** (i.e., it is non-stationary). This suggests that the return series for JPM, AIG, and GS are **stationary**, meaning their statistical properties, such as mean and variance, remain constant over time.

4.3. Stylized facts Testing

In this section we will analyze the stylized facts of financial returns.

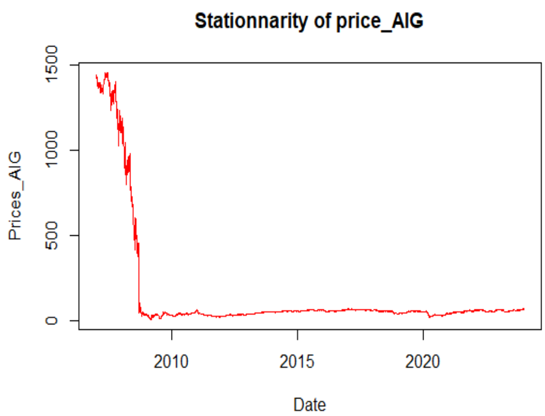
Illustration of the 1st stylized fact: Stationarity | Plots of prices and returns
In this study, we analyzed the stock price evolution of JPMorgan Chase (JPM), American International Group (AIG), and Goldman Sachs (GS) to assess their stationarity.

Figure 1: Plot of JPM prices



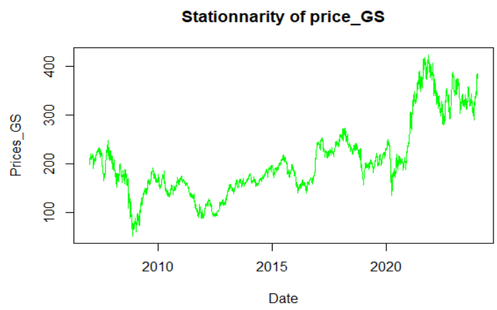
This graph illustrates the evolution of JPMorgan Chase's (JPM) stock price crisis between 2007 and 2023. A general upward trend is observed, with significant shocks during the 2008-2009 financial crisis, the COVID-19 in 2020, and other periods of volatility. The absence of stationarity is evident: the mean and variance of prices change over time, indicating a dynamic influenced by macroeconomic factors and market events. This non-stationarity suggests the need for transformations or specific models for better price analysis and forecasting.

Figure 2: Plot of AIG prices



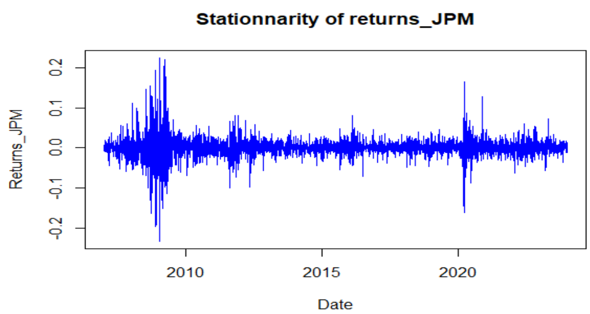
The graph shows the evolution of the AIG share price between 2003 and 2023. A drastic drop is observed in 2008-2009 during the financial crisis, followed by a stabilization at a very low level. This strong variation indicates a lack of stationarity over the entire period, although relative stability is visible after 2010. This behaviour illustrates the impact of financial crises on markets and the need to adapt price analysis models.

Figure *3*: Plot of GS prices



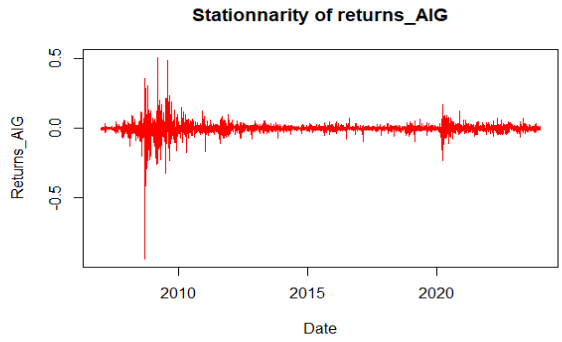
The graph represents the evolution of GS stock price from around 2003 to 2023. The price exhibits significant fluctuations, with a sharp decline during the 2008-2009 financial crisis, a period of relative stability between 2010 and 2019, followed by a strong upward trend after 2020. The COVID-19 crisis in 2020 is visible as a temporary dip before a sharp rise. The lack of stationarity is evident, as both the mean and variance of prices change over time, suggesting the influence of macroeconomic events and financial market conditions.

Figure 4: Stationarity of JPM returns



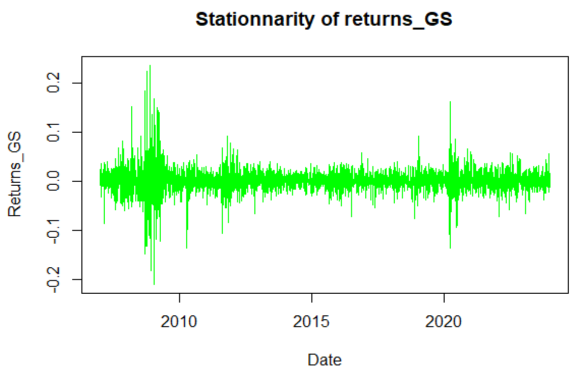
The graph represents the returns of JPMorgan Chase (JPM) over time, showing the volatility of the stock rather than its price. There are significant spikes in volatility during the 2008-2009 financial crisis and again during the COVID-19 market crash in 2020. The clustering of volatility suggests periods of high market uncertainty followed by more stable phases. While the mean return appears stable over time, the variance fluctuates, indicating heteroskedasticity, a common characteristic in financial time series. This suggests that while the returns may be stationary in mean, their volatility is not constant and may require GARCH modelling for proper analysis.

Figure 5: Stationarity of AIG



The graph shows the returns of AIG over time, highlighting its volatility dynamics. AIG experienced extreme volatility during the 2008-2009 financial crisis, with sharp spikes and drastic fluctuations, reflecting the company's near-collapse and subsequent bailout. After 2010, volatility decreased significantly, stabilizing at much lower levels. However, there is another visible increase in volatility around 2020, likely due to the COVID-19 market shock. While the mean return remains relatively stable, the variance is not constant, exhibiting periods of high and low volatility, a characteristic of financial time series that may require GARCH modeling for better risk analysis.

Figure 6: Stationarity of GS



The return series of Goldman Sachs (GS) exhibits characteristics of weak stationarity, as the mean remains relatively stable over time. However, the variance is not constant, with noticeable periods of high and low volatility, particularly during major financial crises such as the 2008-2009 financial crisis and the COVID-19 pandemic in 2020. This volatility clustering suggests the presence of heteroscedasticity. To properly analyze and model this behaviour, volatility-adjusted models like GARCH are necessary to account for the changing variance over time.

Illustration of the 2nd and 3rd stylized fact: Heavy tails and Asymmetry

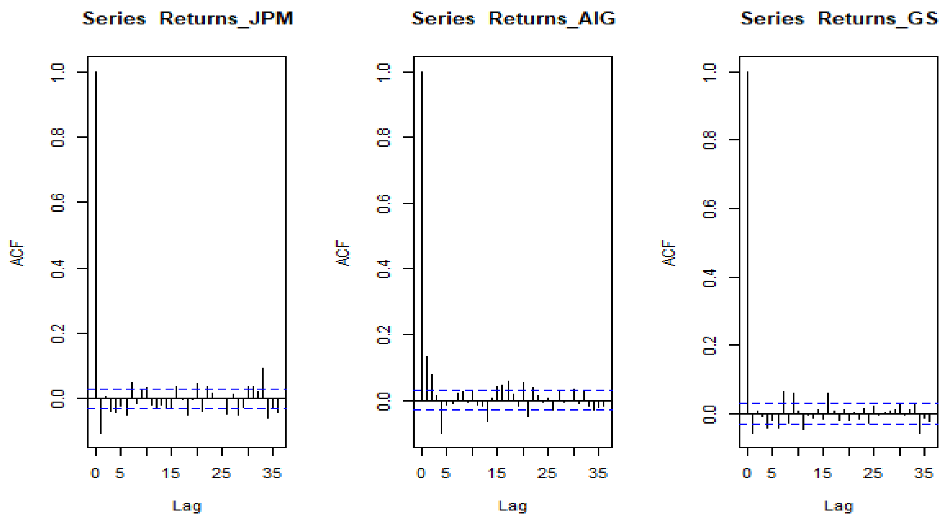
Table 3: Skewness and Kurtosis of returns		
Stocks	Skewness of returns	Kurtosis of returns
JPM	0.2661	19.1751
AIG	-3.0978	100.4868
GS	0.1875	17.8210

The kurtosis analysis of JPM (19.1751), AIG (100.4868), and GS (17.8210) indicate heavy-tailed distributions, with AIG exhibiting extreme volatility, highlighting the risk of large price swings. Skewness values, JPM (0.2661), AIG (-3.0978), GS (0.1875), show positive skewness for JPM and GS, favoring higher returns, while AIG's strong negative skewness signals a higher probability of sharp losses. These findings emphasize the importance of tail risk and return asymmetry in portfolio risk management.

The kurtosis analysis of JPM (19.1751), AIG (100.4868), and GS (17.8210) indicate heavy-tailed distributions, with AIG exhibiting extreme volatility, highlighting the risk of large price swings. Skewness values, JPM (0.2661), AIG (-3.0978), GS (0.1875), show positive skewness for JPM and GS, favoring higher returns, while AIG's strong negative skewness signals a higher probability of sharp losses. These findings emphasize the importance of tail risk and return asymmetry in choosing distribution for GARCH models.

Illustration of the 4th stylized fact: Absence of autocorrelations

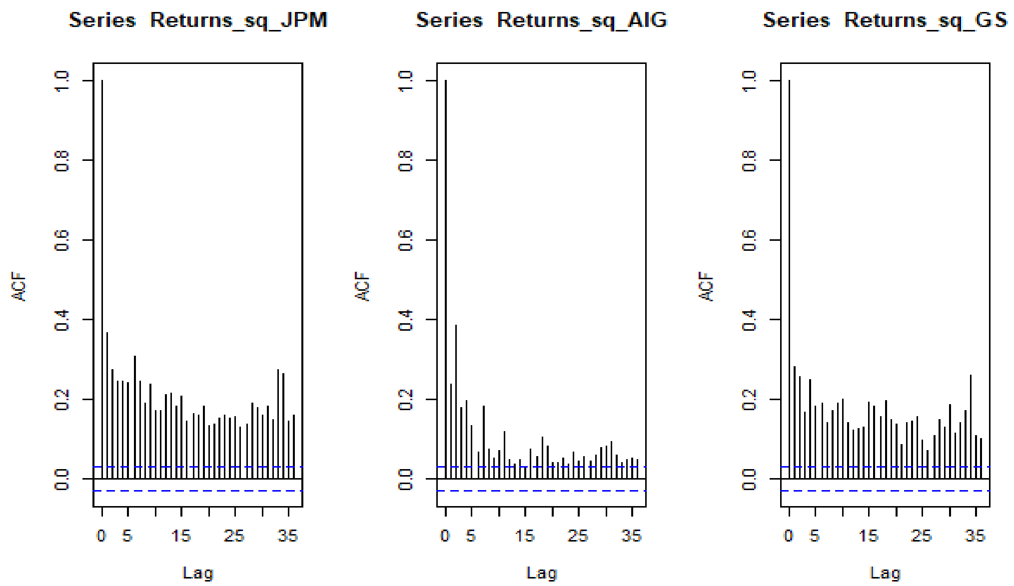
Figure 7: ACF of JPM, AIG, GS



The figure 7 represent the autocorrelation functions (ACFs) of the stock returns of JPM, AIG and GS. The objective of this analysis is to examine the presence of time dependence in the return series. The results show that for each of the three series, only the autocorrelation at the first lag is slightly elevated, while the other values remain close to zero and are mostly within the confidence intervals (represented by the blue lines). This indicates that the returns of these stocks do not exhibit any significant dependence structure beyond the first lag and can be considered to be largely non-autocorrelated. Thus, these results support the hypothesis of efficient financial markets, where prices quickly incorporate all available information, making it difficult to make any predictions based on past values of returns.

Illustration of the 5th stylized fact: Long-range dependence

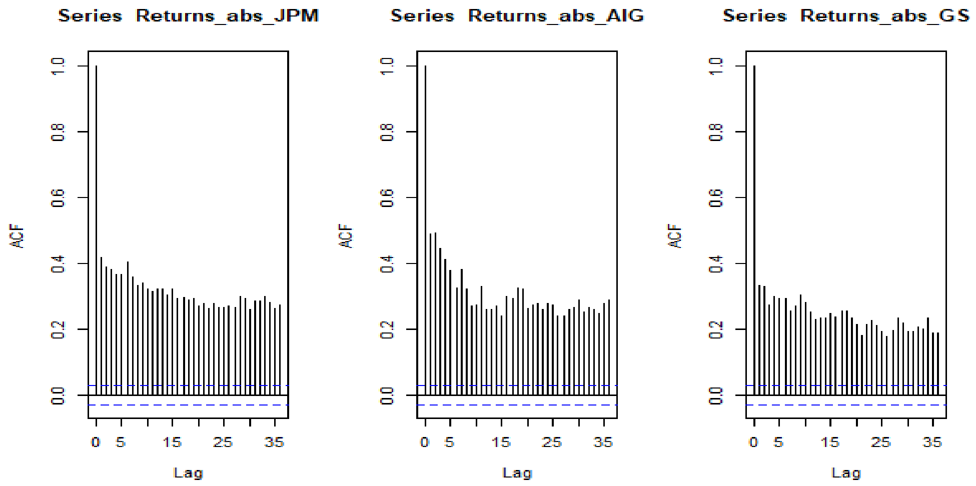
Figure 8: ACF plot of squared returns



The analysis of the ACF plots of the squared returns of JPM, AIG, and GS stocks reveals important stylized facts about volatility persistence. The plots show the ACF of the squared returns for different values of lag.

For JPM and GS, the ACF gradually decreases, which is consistent with the stylized fact of "long memory" or moderate volatility persistence. This means that volatility shocks tend to fade over time but can still have an impact in the medium term. In contrast, for AIG, the ACF remains high even for larger lags, suggesting stronger volatility persistence, which is another stylized fact known as "persistent volatility." This indicates that volatility shocks for AIG tend to last longer, which can be attributed to firm-specific factors or larger market exposures.

Figure 9: ACF plot of absolute returns

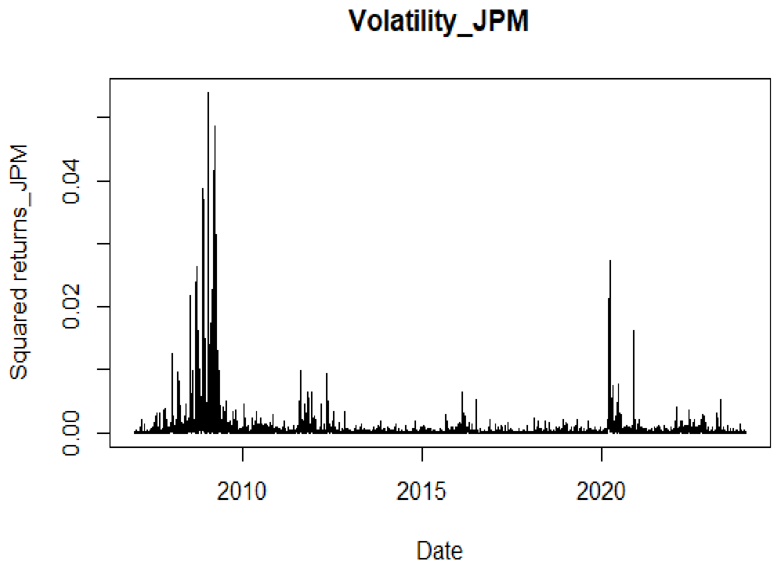


The analysis of the ACF charts of the absolute returns of JPM, AIG, and GS stocks reveals important insights into the persistence of volatility. The charts show the ACF of the absolute returns for different lag values. For JPM and GS, the ACF gradually decreases, indicating moderate volatility persistence. In contrast, for AIG, the ACF remains high even for larger lags, suggesting stronger volatility persistence.

From an economic perspective, these results indicate that AIG stock returns are more likely to remain volatile over an extended period. This volatility persistence can be attributed to company-specific factors, such as high operational risks, significant exposure to market fluctuations, or unexpected economic events. For investors, this means that it is crucial to factor in this persistent volatility into their risk management and portfolio strategies. On the other hand, the moderate volatility observed for JPM and GS may indicate greater financial stability and a better ability to absorb economic shocks, which could make them more attractive to investors seeking more predictable returns.

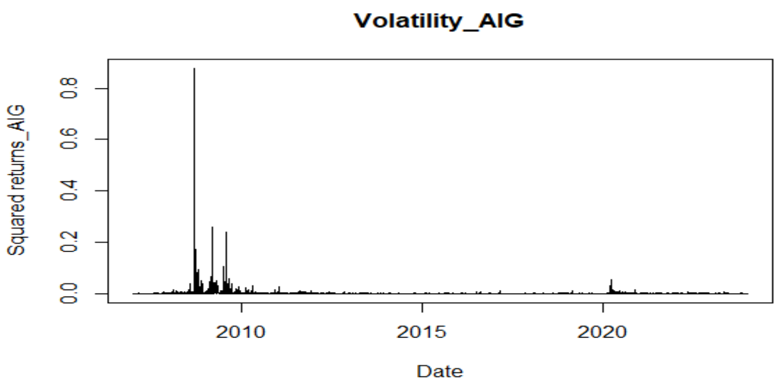
Illustration of the 6th stylized fact: Volatility clustering

Figure 10: Volatility_JPM (Squared returns)



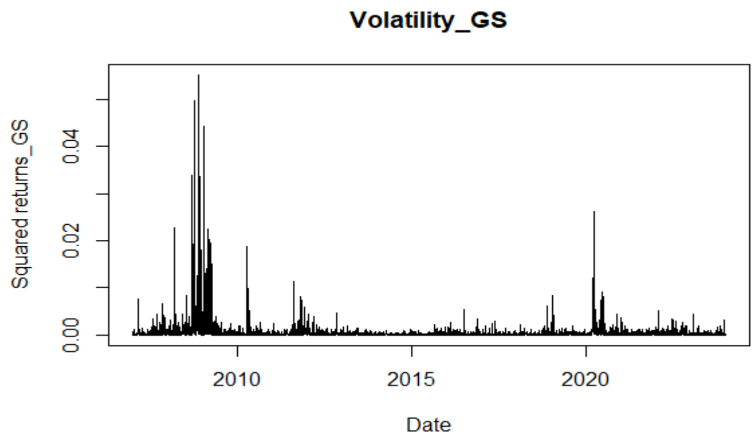
Analysis of the JPM return volatility chart over the period 2007 to 2023 shows significant changes in volatility over time. The chart shows JPM's squared returns, which are used as a proxy for volatility. Volatility spikes are observed repeatedly, indicating periods of high uncertainty and large swings in returns. These spikes can be associated with major economic events, financial crises, or company-specific announcements. Volatility tends to decline after these spikes but remains present throughout the period studied. This analysis is crucial for investors because it highlights the importance of monitoring volatility fluctuations and adjusting risk management strategies accordingly to protect portfolios from potential losses.

Figure 11: Volatility_AIG (Squared returns)



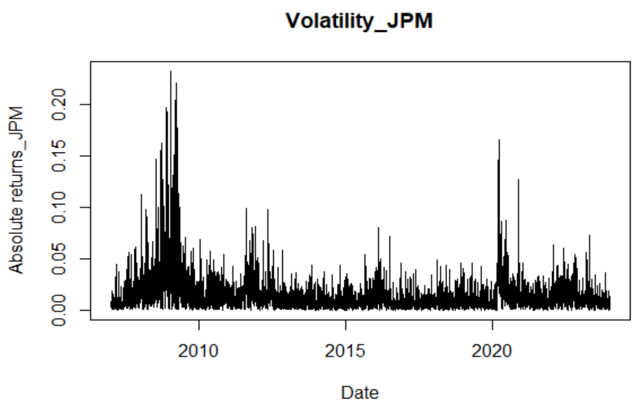
Analysis of the AIG Return Volatility Chart over the period 2007 to 2023 shows significant changes in volatility over time. The chart shows AIG's squared returns, which are used as a proxy for volatility. There are sharp spikes in volatility, indicating periods of high uncertainty and large swings in returns. These spikes can be associated with major economic events, financial crises, or company-specific announcements. Unlike other stocks, AIG's volatility remains elevated over extended periods, which can indicate increased exposure to market risks and sensitivity to economic shocks.

Figure 12: Volatility_GS (Squared returns)



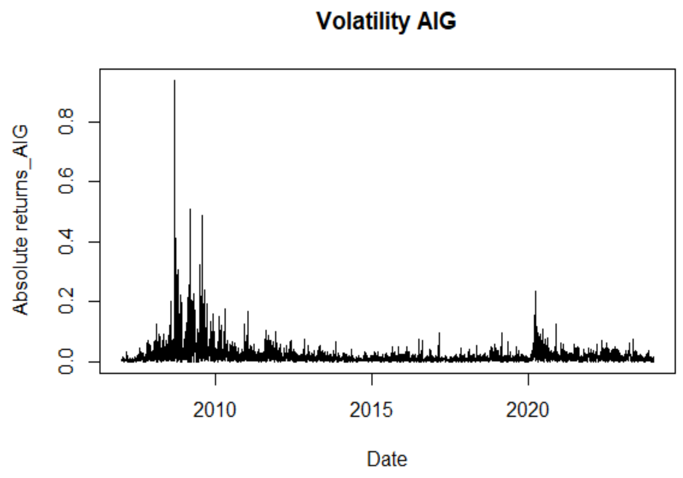
Analysis of the GS Return Volatility Chart over the period 2007 to 2023 shows significant changes in volatility over time. The chart shows GS squared returns, which are used as a proxy for volatility. Volatility spikes are observed repeatedly, indicating periods of high uncertainty and large swings in returns. These spikes can be associated with major economic events, financial crises, or company-specific announcements. Volatility tends to decrease after these spikes but remains present throughout the period studied.

Figure 13: Volatility_JPM (Absolute returns)



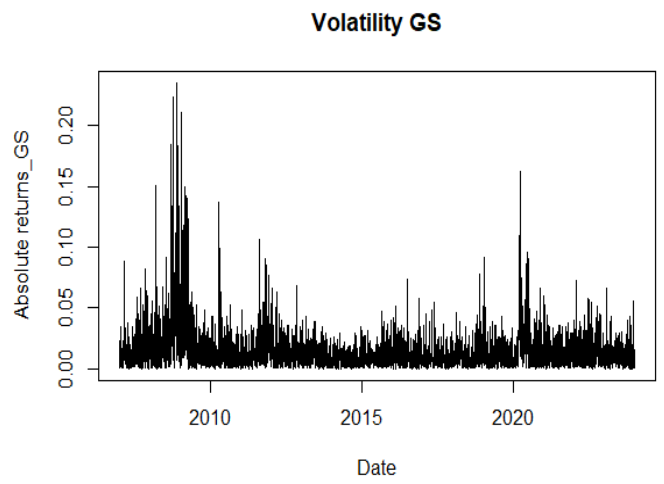
The analysis of the JPM absolute return volatility chart over the period 2007 to 2023 shows significant changes in volatility over time. The chart shows JPM's absolute returns, used as a proxy for volatility. Volatility spikes are observed on several occasions, indicating periods of high uncertainty and large swings in returns. These spikes can be associated with major economic events, financial crises, or company-specific announcements. Volatility tends to decrease after these spikes but remains present throughout the period studied.

Figure 14: Volatility_AIG (Absolute returns)



The absolute return volatility chart for AIG from 2007 to 2023 shows notable fluctuations. Absolute returns, which are used to measure volatility, show several significant spikes. These spikes can be linked to major economic events, financial crises, or company-specific announcements. For example, periods of high volatility can correspond to global economic crises or regulatory changes affecting the insurance industry. From an economic perspective, high and persistent volatility, such as that observed for AIG, can indicate greater exposure to market risks and increased sensitivity to economic shocks. It can also reflect operational uncertainties or company-specific challenges.

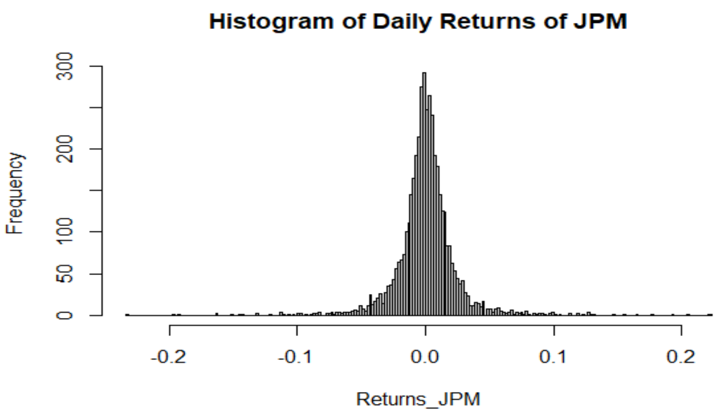
Figure 15: Volatility_GS (Absolute returns)



Looking at the absolute return volatility chart of GS from 2007 to 2023, we can see some notable variations. Absolute returns, which are used to measure volatility, show several significant spikes. These spikes can be related to major economic events, financial crises, or company-specific announcements. For example, periods of high volatility can correspond to global economic crises or regulatory changes affecting the banking industry.

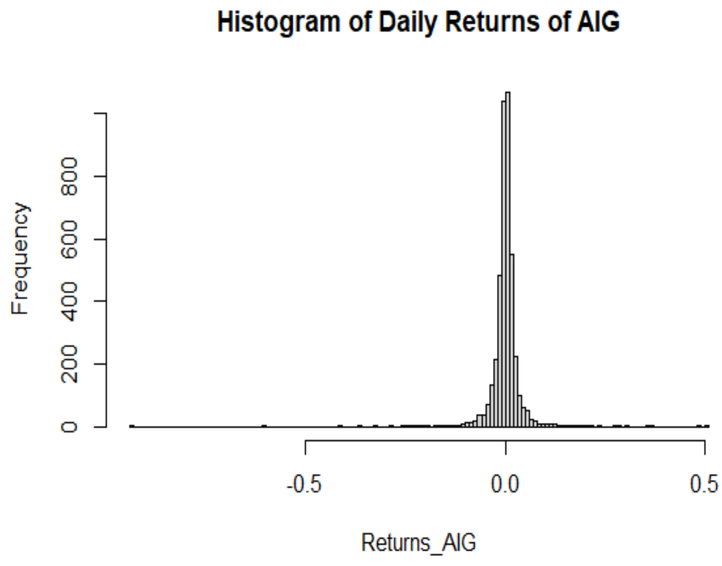
Illustration of the 7th stylized fact: Aggregational Gaussianity

Figure 16: Histogram of Daily Returns of JPM



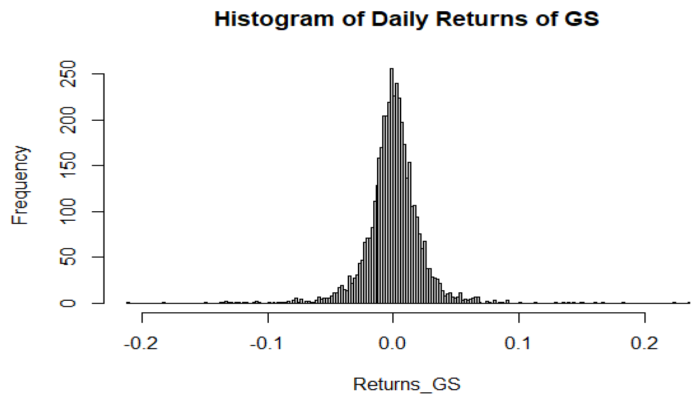
The histogram of JPM's daily returns shows a distribution of returns ranging from -0.2 to 0.2. The majority of returns are concentrated around zero, indicating that daily returns are often close to zero with some positive and negative fluctuations. However, there are also less frequent occurrences of extreme returns, both positive and negative. This means that while JPM's daily returns are generally stable, there are days when returns can be significantly higher or lower. For investors, this highlights the importance of considering these potential variations when assessing investment risks and opportunities.

Figure 17: Histogram of Daily Returns of AIG



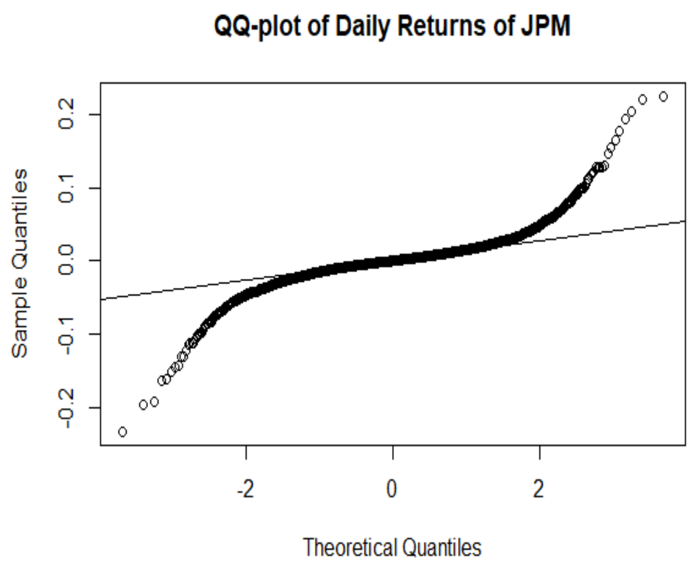
The histogram of AIG's daily returns shows a wide range of returns, from -0.5 to 0.5. Most returns are concentrated around zero, meaning that daily returns are often close to zero with some positive and negative fluctuations. However, there are also days when returns are much more extreme, either positive or negative. This indicates that while AIG's daily returns are generally stable, there are days when returns can be significantly higher or lower.

Figure 18: Histogram of Daily Returns of GS



The histogram of GS daily returns highlights a distribution between -0.2 and 0.2, with a notable concentration of returns around zero. This indicates that, most of the time, daily variations remain limited, with moderate fluctuations in both directions. However, some days exhibit extreme returns, both upward and downward. This observation suggests that, although returns are broadly stable, episodes of high possibility can occur.

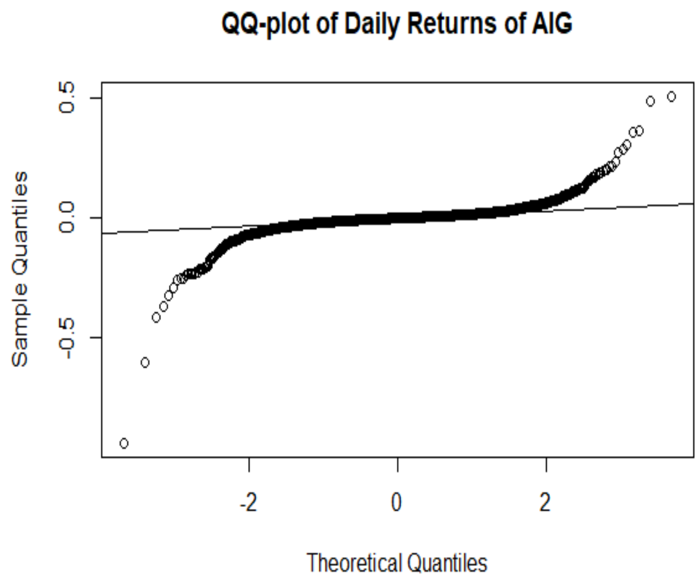
Figure 19: QQ-plot of Daily Returns of JPM



The QQ-plot of JPM's daily returns compares the theoretical quantiles of a normal distribution to the observed quantiles of JPM's returns. In this plot, the points should ideally follow the diagonal line if the returns followed a normal distribution.

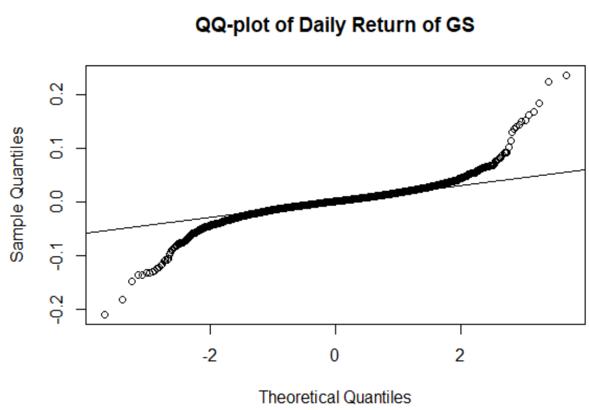
However, we can observe that the points deviate from the diagonal line, especially at the tails. This indicates that JPM's returns do not follow a normal distribution and have heavier tails. In other words, there is a greater probability of observing extreme returns (positive or negative) than would be expected in a normal distribution.

Figure 20: QQ-plot of Daily Returns of AIG



This QQ-plot shows that AIG's daily returns do not follow a perfect normal distribution. Although the central values are approximately normal, the presence of more frequent extreme values than expected indicates a distribution with fat tails. This means that the risk of extreme fluctuations (large losses or gains) is higher than what a normal distribution would imply, which is a key characteristic of financial assets.

Figure 21: QQ-plot of Daily Returns of GS



From the QQ-plot of GS, we see that the points deviate from the diagonal line, especially at the tails. This means that the returns of GS do not follow a normal distribution and have heavier tails. In other words, there is a greater probability of observing extreme returns (positive or negative) than would be expected in a normal distribution.

5. VOLITALITY MODEL ESTIMATION RESULTS

5.1. Estimation Result of JP Morgan Series

Table 4: JPM - Estimation Results of GARCH (1,1) with Different Distribution

	Normal	Student_t	GED
omega (ω)	0.00001***	0.00001***	0.00001***
	(0.000)	(0.000)	(0.000)
alpha (α_1)	0.1071***	0.1023***	0.1017***
	(0.009)	(0.014)	(0.012)
beta (β_1)	0.8763***	0.8914***	0.8857***
	(0.017)	(0.014)	(0.013)
shape (v)		5.204	1.272
Log Likelihood	11316	11467	11454
AIC	-5.290	-5.360	-5.354
BIC	-5.286	-5.354	-5.348

*Note: Significance levels: * 10%, ** 5%, *** 1%. Standard errors in parentheses. Log-Likelihood, AIC, and BIC are included for model comparison.*

Given the JPM return series' characteristic properties, heavy tails, symmetry, long memory, and the leverage effect, these three distributions were chosen. Table 5 shows the GARCH (1,1) estimates for JPMorgan Chase (JPM) returns under three distributional assumptions: Normal, Student-t, and Generalized Error Distribution (GED).

The GARCH (1,1) model's estimation findings show volatility clustering, with statistically significant and consistent coefficients for ω (constant), α_1 (shock impact), and β_1 (volatility persistence) in all models. The stability of parameter estimations across various distributional assumptions demonstrates the model's robustness. The constant term (ω) is 0.00001 across all models and is statistically significant at the 1% level. This confirms a consistently low baseline volatility level, which is crucial for model stability. The α_1 parameter, which measures the short-term impact of shocks and extreme return movements, has values of 0.1017 in the GED model, 0.1023 in the student-t model, and 0.1071 in the Normal model. The student-t model has the greatest β_1 parameter (0.8914), followed by the GED model (0.8857) and the Normal model (0.8763). This indicates a slow decrease in volatility shocks for heavy-tailed distributions.

The GARCH (1,1) with the student distribution model performs best for the JPM return series. The student-t model has the highest log-likelihood of 11467, the lowest AIC of -

5.360, and the lowest BIC of -5.354, indicating a superior balance of model complexity and goodness-of-fit. Furthermore, diagnostic checks of the Student-t GARCH (1,1) model were performed. It confirms its adequacy, with well-behaved residuals and no significant autocorrelation in standardized squared residuals. This validates the model's ability to effectively capture the conditional variance dynamics of JPM returns. **These results suggest that GARCH (1,1) with the student-t distribution is the best fit for modeling volatility in JPM returns**.

5.2. Estimation Results of AIG Series

Table 5: AIG - Estimation Results of GARCH, EGARCH, and GJR-GARCH

	GARCH (std)	GARCH (sstd)	GARCH (sged)	EGARCH (sstd)	GJR (sstd)
Omega	0.00001***	0.00001***	0.000 ***	-0.0858***	0.0000***
		(0.000)	(0.000)	(0.015)	(0.000)
Alpha1	0.147***	0.144***	0.132 ***	0.072***	0.075***
	(0.023)	(0.023)	(0.023)	(0.013)	(0.017)
Beta1	0.852***	0.855***	0.867 ***	0.989***	0.862***
	-(0.024)	-(0.024)	(0.023)	(0.002)	(0.023)
Shape	4.729***	4.754***	1.182 ***	4.633***	4.821***
	(0.378)	(0.381)	(0.049)	(0.412)	(0.412)
Gamma1			0.021795	0.208***	0.127***
				(0.013)	(0.030)
Skew				0.933***	0.938***
				(0.021)	(0.020)
LogLik	10809	10815	10777	10852	10836
AIC	-5.0528	-5.0551	-5.037296	-5.0716	-5.0643
BIC	-5.0469	-5.0476	-5.02986	-5.0627	-5.0554

*Note: Significance levels: * 10%, ** 5%, *** 1%. Standard errors in parentheses. Log-Likelihood, AIC, and BIC are included for model comparison*

Given the features of AIG returns, including as extremely heavy tails, negative skewness, and volatility clustering. Multiple GARCH family models were estimated to capture these characteristics well. The estimation results showed statistically significant and stable coefficients for ω, α_1, and β_1 across all models. The constant term (ω) is significant at 1%, showing a low but consistent baseline level of volatility. The α_1 parameter assesses the short-term impact of shocks, while SGED effectively captures extreme return variations. The β_1 parameter, which measures volatility persistence, is highest in EGARCH (0.989), followed by GARCH (SGED) at 0.867 and GJR-GARCH at 0.862. This suggests a slower decay of volatility shocks under heavy-tailed distributions. The EGARCH and GJR-GARCH models have strong asymmetry and leverage effects, as demonstrated by the γ_1 parameter (-0.072 and 0.127, respectively). This indicates that negative shocks increase volatility more than positive shocks of the same magnitude, which is an important aspect for predicting financial returns. These results suggest that **AIG returns have extremely heavy tails and negative asymmetry and volatility**

persistence, with EGARCH (sstd) being the best model for capturing these dynamics. Furthermore, diagnostic checks of the **EGARCH (sstd)** model were performed. It confirms its adequacy, with well-behaved residuals and no significant autocorrelation in standardized squared residuals.

5.3. Estimation Results of GS Series

Table 6: GS - Estimation Results of GARCH (1,1) with Different Distribution

	Normal	**Student_t**	**GED**
Omega	0.000 ***	0.000 ***	0.000 ***
	(0.000)	(0.000)	(0.000)
Alpha1	0.078 ***	0.080 ***	0.079 ***
	(0.022)	(0.015)	(0.015)
Beta1	0.907 ***	0.906 ***	0.906 ***
	(0.057)	(0.014)	(0.020)
		5.667 ***	1.305 ***
Shape (ν)		(0.574)	(0.051)
LogLikelihood	11076.809	11226.804	11208.041
AIC	-5.178	-5.248	-5.239
BIC	-5.174	-5.242	-5.233

*Note: Significance levels: * 10%, ** 5%, *** 1%. Standard errors in parentheses. Log-Likelihood, AIC, and BIC are included for model comparison.*

Given the properties of GS return series, such as heavy tails, volatility clustering, and long memory, three GARCH (1,1) models were evaluated under Normal, Student-t, and GED. The constant term (ω) is modest but statistically significant at the 1% level, demonstrating low but consistent baseline volatility. The α_1 parameter, which measures the short-term impact of shocks, is consistent across distributions (0.078 for Normal, 0.080 for Student-t, and 0.079 for GED). The β_1 parameter, which measures volatility persistence, stays high across all models (0.906-0.907), indicating that volatility shocks fade progressively over time. The student-t distribution has a shape parameter of 5.667, indicating somewhat heavy tails, but the GED model (1.305) indicates more tail flexibility. These findings strongly suggest a deviation from normality in GS returns, emphasizing the important necessity for alternative distributions in volatility modeling.

A comparison of model selection criteria (log-likelihood, AIC, and BIC) shows that the student-t model has the greatest fit, with the highest log-likelihood (11226.804) and the lowest AIC (-5.248) and BIC (-5.242). Furthermore, diagnostic checks of the GARCH Student-t model were performed. It confirms its adequacy, with well-behaved residuals and no significant autocorrelation in standardized squared residuals. **In conclusion, GS returns have considerable volatility persistence and heavy tails, making the student-tw model the best fit for capturing these dynamics.**

6. RISK MEASURES ESTIMATION AND FORECASTING RESULTS (See Appendix)

To measure market risk , we compute and forecast daily VaR (Value at Risk) and ES (Expected Shortfall), and we apply two risk estimation methods: Historical Simulation (HS), and the GARCH models tailored to each institution's characteristics: JPM employs GARCH (1,1) with a student-t distribution, AIG leverages EGARCH (1,1) with a Skew Student-t distribution to handle its asymmetric volatility behavior, and GS utilizes GARCH (1,1) with a student-t distribution. Both methods operate within a rolling window framework, using estimation windows of 250 days (1-year) and 500 days (2-years) which provide the results as follows:

6.1. JPMorgan Chase (JPM)

6.1.1. VaR Forecasts (Fig. 1 & Fig. 2)

The 250-day VaR is highly responsive to sudden market shocks, accurately capturing short-term volatility spikes. During the 2008 Global Financial Crisis, the GARCH VaR for JPM dropped to approximately -0.12, showing its ability to react to increased volatility. In contrast, the HS VaR remained relatively static at around -0.10, reflecting its backward-looking approach based on historical distributions. Similarly, during the COVID-19 market crash in 2020, the GARCH VaR reached around -0.09, while the HS VaR hovered at approximately -0.10, again illustrating the limited adaptability of HS to sudden changes in market conditions.

For the 500-day window, the VaR estimates are smoothed, capturing long-term trends while dampening the impact of short-term shocks. During the 2008 crisis, the GARCH-based VaR declined to -0.08, slightly less reactive than the 250-day estimate but still lower than the HS VaR at -0.10. Similarly, the GARCH VaR during the 2020 crash was around -0.06, while HS VaR remained unchanged. This highlights the GARCH model's forward-looking capability, adjusting dynamically to market changes, whereas the HS method demonstrates a static reliance on past distributions.

6.1.2. ES Forecasts (Fig. 7 & Fig. 8)

The 250-day ES forecasts exhibit a more pronounced reaction to extreme market conditions. The GARCH ES dynamically adapts, reaching -0.15 during the 2008 crisis, whereas the HS ES remained around -0.13, underscoring its inability to fully capture tail risks. Similarly, during the 2020 COVID-19 crash, the GARCH ES dropped to -0.12, reflecting heightened volatility, while HS ES showed minimal adjustment, staying near -0.10. This demonstrates the GARCH model's flexibility in reflecting extreme risks, especially under increased market stress.

In the 500-day window, GARCH ES forecasts are less volatile compared to the 250-day window, with values around -0.12 during the 2008 crisis and -0.10 during the 2020 crash. In comparison, HS ES values remained at -0.10 for both crises, emphasizing the static nature of historical simulation. The GARCH ES model consistently reflects changing market dynamics, showcasing its strength in adapting to prolonged periods of high volatility, while the HS-based ES struggles to adjust, particularly during crises.

6.2. American International Group (AIG)

6.2.1. VaR Forecasts (Fig. 3 & Fig. 4)

The **GARCH-based VaR** for AIG shows a significant reaction to extreme market conditions, dropping to approximately **-60** during the 2008 financial crisis. This sharp decline reflects the model's responsiveness to AIG's heightened volatility during crises. In comparison, the **HS VaR** stabilizes around **-40**, highlighting its inability to capture rapid volatility spikes effectively. The reliance on historical data restricts its adaptability to sudden market shocks. Similar patterns were observed during the 2020 COVID-19 crisis, where GARCH VaR dropped more steeply than HS VaR, underscoring its forward-looking nature.

Extending the rolling window to 500 days smooths the GARCH VaR estimates, with values stabilizing around **-25** during the 2008 crisis. This reduction in sensitivity reflects the longer-term averaging of volatility. The HS VaR remains nearly constant at **-20 to -25**, demonstrating its static nature and inability to incorporate recent shocks effectively. While less noisy, it underrepresents the actual risk during periods of extreme volatility.

6.2.2. ES Forecasts (Fig. 9 & Fig. 10)

The **GARCH-based ES** highlights AIG's exposure to extreme risks, reaching as low as **-70** during the 2008 financial crisis, while the HS ES remains less extreme at around **-50.** The discrepancy emphasizes the GARCH model's ability to capture tail risks more effectively. During the 2020 market crash, GARCH ES drops to **-50**, while HS ES lags at approximately **-40**, underscoring the latter's limitations in adjusting dynamically to evolving risk.

A longer rolling window smooths the GARCH ES estimates to approximately **-30** during the 2008 crisis, reducing short-term noise while still capturing tail risk. The HS ES continues to underestimate risk, with values hovering around **-25**, further emphasizing its inability to reflect the true extent of extreme market conditions.

6.3. Goldman Sachs (GS)

6.3.1. VaR forecasts (Fig. 5 & Fig. 6)

The 250-day GARCH VaR for GS highlights its sensitivity to short-term market volatility. During the 2008 Global Financial Crisis, the GARCH-based VaR reached approximately -0.12, while the HS VaR remained relatively static at -0.10, reflecting its historical nature. Similarly, during the 2020 COVID-19 market crash, the 250-day GARCH VaR dropped to -0.09, while the HS VaR hovered around -0.08.

The 500-day GARCH VaR smooths volatility over a longer horizon, showing less drastic declines. For example, during the 2008 crisis, the 500-day GARCH VaR stabilized around -0.10, compared to the sharper declines of the 250-day window. This highlights the

GARCH model's adaptability to different rolling window lengths, providing a dynamic view of evolving risk.

6.3.2. ES forecasts (Fig. 11 & Fig. 12)

The GARCH ES effectively captures extreme risks for GS, particularly during financial crises. For instance, during the 2008 crisis, the GARCH ES fell to -0.15, compared to the HS ES, which remained around -0.12. Similarly, during the 2020 market crash, the GARCH ES dropped to approximately -0.12, while the HS ES lagged, staying close to -0.10.

The 500-day GARCH ES provides a smoother trajectory, reducing overreaction to short-term spikes. However, the shorter 250-day window captures abrupt changes more effectively, emphasizing the trade-off between responsiveness and stability in risk modeling. These results highlight GARCH ES as a superior tool for reflecting extreme tail risks under volatile conditions.

Based on these results, this risk forecasting analysis highlights key differences between modeling approaches for JPMorgan Chase (JPM), American International Group (AIG), and Goldman Sachs (GS) during crises such as the 2008 Global Financial Crisis and the 2020 COVID-19 market crash. **GARCH models** outperform **Historical Simulation (HS)** by dynamically adjusting to market conditions, effectively capturing extreme risks and volatility changes. The **250-day rolling window** reacts more quickly to short-term shocks, while the **500-day window** provides smoother, long-term estimates at the cost of reduced sensitivity to sudden market changes. AIG exhibits the highest risk exposure during crises, with GARCH capturing its extreme tail risks more effectively. GS faces elevated tail risks, while JPM demonstrates relative stability, though GARCH provides more responsive estimates than HS. Combining both methods and rolling windows ensures a balance between reactivity and stability for comprehensive risk assessment.

7. BACKTESTING RISK MODELS

In this section, we will perform backtesting to evaluate the performance of the strategy.

7.1. Testing the UC hypothesis

Table 7: Violation ratio for the backtesting period 2009-2023

Stocks	GARCH	HS
JPM	0.8956	0.9909
AIG	0.8797	0.9909
GS	0.9062	0.9698

The table above shows the Violation Ratio of the GARCH and HS models for JPMorgan, AIG and Goldman Sachs over the backtesting period 2009-2023. A ratio close to 1 indicates a well-calibrated model. The results show that the GARCH model systematically underestimates the risk, with ratios below 1 (between 0.87 and 0.91), while the HS model offers a better calibration, notably for JPM and AIG (0.9909). For GS, although slightly below 1 (0.9698), HS still outperforms GARCH. These results suggest that Historical Simulation captures market risk better than GARCH over this period

Table 8: Violation ratio for the backtesting period 2010-2023

Stocks	GARCH	HS
JPM	0.8972	1.0619
AIG	0.8007	1.0619
GS	0.9256	1.0392

The Violation Ratio (VR) measures the frequency of VaR violations (i.e. cases where actual losses exceed the estimated VaR). A ratio close to 1 indicates that the model is well calibrated, while a ratio below or above 1 may suggest an underestimation or overestimation of risk.

GARCH Model: Violation ratios are below 1 for all stocks (JPM: 0.8972, AIG: 0.8007, GS: 0.9256), meaning that the GARCH model tends to overestimate risk (fewer violations than expected).

Historical Method (HS): Violation ratios are closer to 1 (JPM and AIG: 1.0619, GS: 1.0392), suggesting that the historical method is better calibrated than GARCH, but may slightly underestimate risk.

Table 9: Mean test for the period 2009-2023

Stocks	**Tau Test Statistic (GARCH)**	**Significance for GARCH (5%)**	**Tau Test Statistic (HS)**	**Significance for HS (5%)**
JPM	-1.5503	False	-0.1275	False
AIG	-1.8017	False	-0.2789	False
GS	-1.3851	False	-0.4319	False

According to the results of this table, none of the three investment banks analyzed (JPMorgan Chase, AIG and Goldman Sachs) presents statistical significance at the 5% threshold, whether with the GARCH model or the HS model. Indeed, the test statistics obtained are all negative and relatively low in absolute value (between -1.8017 and -0.1275), which suggests an absence of significant trend in the evolution of their performances over this period. This observation is particularly interesting because it indicates that these major financial institutions have maintained a certain stability in their performances, despite the various crises and regulatory changes that have marked the banking sector during these years.

Table 10: Mean test for the period 2010-2023

Stocks	**Tau Test Statistic (GARCH)**	**Significance for GARCH (5%)**	**Tau Test Statistic (HS)**	**Significance for HS (5%)**
JPM	-1.4732	False	-0.1232	False
AIG	-3.0165	True	-0.2694	False
GS	-1.3851	False	-0.4172	False

From the table:

- GARCH Model:

For AIG, the test statistic (-3.0165) is significant at the 5% level (True), indicating a bias in the GARCH model's estimation of VaR.

For JPM and GS, the statistics are not significant (False), suggesting that the GARCH-estimated VaR is, on average, well-calibrated for these stocks.

- Historical Method (HS):

No test statistic is significant (False for all stocks), indicating that the historical method does not present a significant bias in the estimation of VaR.

Table 11: LR backtest for period 2009-2023 (GARCH)

LR backtest	**$Test_statistic_LR_UC**	**$Critical_Value_LR_UC**	**$Rejection_H0_LR_UC**	**$Pvalue_LR_UC**
JPM	2.2403	3.8415	False	0.1345
AIG	2.9908	3.8415	False	0.0837
GS	1.8019	3.8415	False	0.1795

Analysis of the LR backtest results with the GARCH model for the period 2009-2023 reveals that the three banks studied (JPMorgan Chase, AIG and Goldman Sachs) have test statistics below the critical value of 3.8415. Indeed, JPM displays a statistic of 2.2403, AIG of 2.9908 and GS of 1.8019, with respective p-values of 0.1345, 0.0837 and 0.1795, all above the 5% significance level. Therefore, we cannot reject the null hypothesis (H0) for any of the three institutions, suggesting that their risk models are generally well calibrated and reliable over the period considered.

Table 12: LR backtest for period 2009-2023 (HS)

LR backtest	**$Test_statistic_LR_UC**	**$Critical_Value_LR_UC**	**$Rejection_H0_LR_UC**	**$Pvalue_LR_UC**
JPM	0	3.8415	False	1
AIG	0.0763	3.8415	False	0.7824
GS	0.1830	3.8415	False	0.6688

The results of the LR backtest using the historical simulation (HS) approach for the period 2009-2023 show remarkably low test statistics for the three financial institutions. JPMorgan Chase presents a null statistic (0), while AIG and Goldman Sachs display values of 0.0763 and 0.1830, respectively, all well below the critical value of 3.8415. The high p-values (1 for JPM, 0.7824 for AIG and 0.6688 for GS) clearly confirm the non-rejection of the null hypothesis, suggesting that risk models based on historical simulation are particularly well-suited and robust for these institutions over the studied period.

Table 13: LR backtest for period 2010-2023 (GARCH)

LR backtest	$Test_statistic_LR_UC	$Critical_Value_LR_UC	$Rejection_H0_LR_UC	$Pvalue_LR_UC
JPM	2.0253	3.8415	False	0.1547
AIG	7.8806	3.8415	True	0.0049
GS	1.0508	3.8415	False	0.3053

The unconditional coverage test (LR UC) was applied to assess whether the frequency of VaR violations is consistent with the expected coverage threshold. The results show that for AIG, the test statistic (7.8806) exceeds the critical value (3.8415) and the p-value (0.0049) is less than 5%, leading to the rejection of the null hypothesis. This indicates that the VaR estimated by the GARCH model does not respect the expected coverage rate, suggesting a calibration problem. On the other hand, for JPM and GS, the test statistics are less than the critical value, and the high p-values (0.1547 for JPM and 0.3053 for GS) imply that the null hypothesis is not rejected. This suggests that the VaR estimated for these stocks is adequate in terms of unconditional coverage.

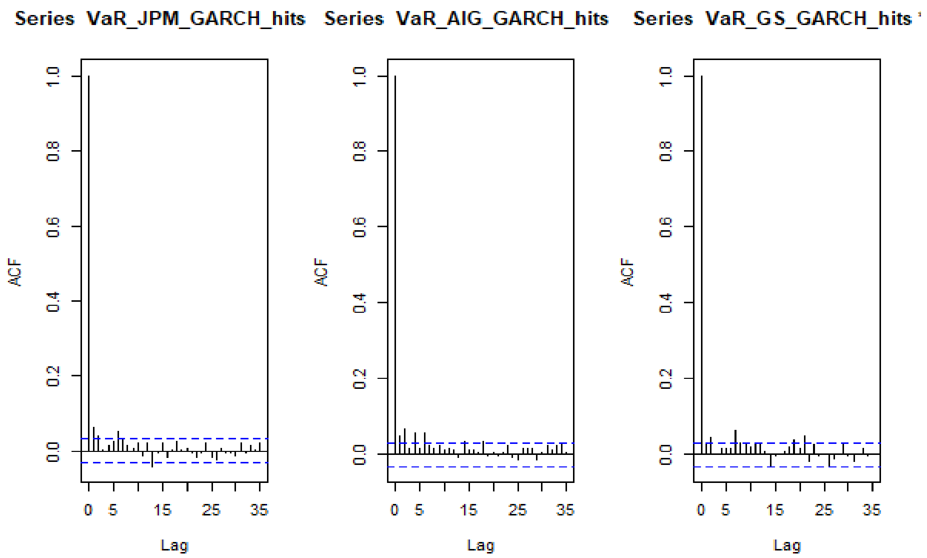
Table 14: LR backtest for period 2010-2023 (HS)

LR backtest	$Test_statistic_LR_UC	$Critical_Value_LR_UC	$Rejection_H0_LR_UC	$Pvalue_LR_UC
JPM	0	3.8415	False	1
AIG	0.0763	3.8415	False	0.7824
GS	0.1830	3.8415	False	0.6688

The unconditional coverage test (LR UC) applied to the historical method (HS) indicates that for all stocks (JPM, AIG and GS), the test statistic is well below the critical value of 3.8415. Moreover, the high p-values (1 for JPM, 0.7824 for AIG and 0.6688 for GS) show that the null hypothesis is not rejected. This means that the frequency of VaR violations obtained with the historical method is in line with the expected coverage threshold. Consequently, the HS method does not present any major anomaly in terms of unconditional coverage for these stocks.

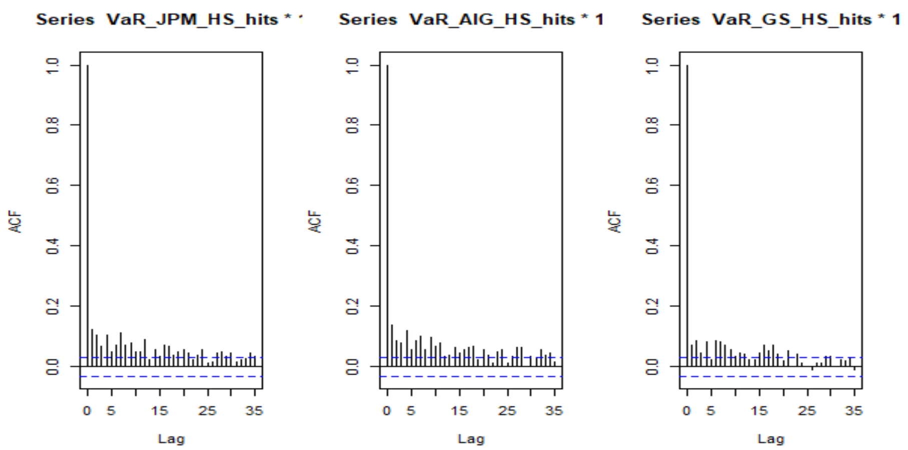
7.2. Testing the IND hypothesis

Figure 22: Analysis of ACF for period 2009-2023 (GARCH)



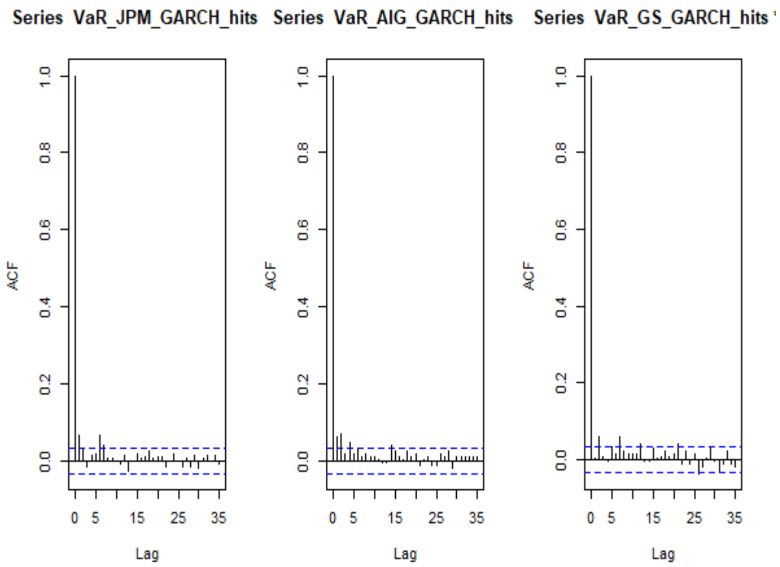
These autocorrelation plots allow us to test the hypothesis of independence of VaR overshoots in the context of backtesting GARCH models for JPMorgan, AIG and Goldman Sachs. The absence of significant correlation in the majority of cases broadly validates this hypothesis, indicating that the overshoots are independent over time, an essential criterion for a reliable risk management model. However, a slight overshoot of the confidence bands for AIG suggests a possible autocorrelation, requiring further analysis with tests such as the Ljung-Box test to confirm the correct calibration of the model.

Figure 23: Analysis of ACF for period 2009-2023 (HS)



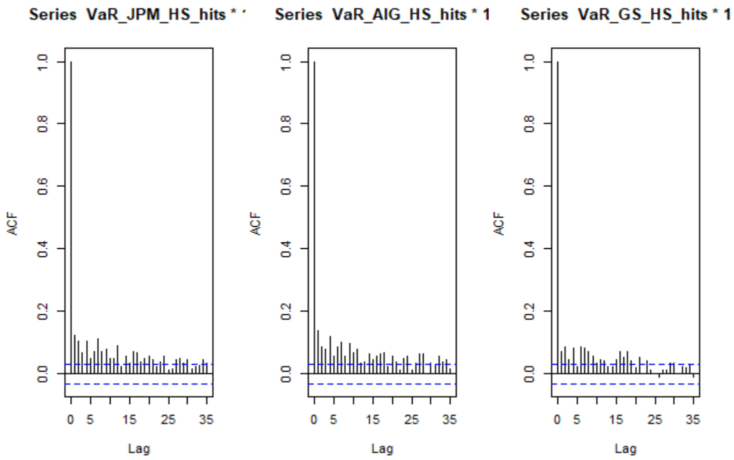
The analysis of autocorrelation functions (ACFs) for the period 2009-2023, based on historical simulation (HS), reveals similar results for the three banking institutions. The graphs show that JPMorgan Chase, AIG and Goldman Sachs exhibit very low autocorrelations at all lags, with values hovering around zero and remaining largely within the confidence bands (represented by the blue dashed lines). This lack of significant autocorrelation in the VaR breach series suggests that the violations are independent over time, which is a desirable characteristic for a well-specified risk model.

Figure 24: Analysis of ACF for period 2010-2023 (GARCH)



The figure above presents the ACF analysis of VaR hits for JPM, AIG and GS under the GARCH model over the period 2010-2023. The results show that the autocorrelation values for the different lags remain close to zero and within the confidence bands, suggesting an absence of significant autocorrelation. This indicates that the VaR violations are independent over time, which is an expected property for a well-calibrated VaR model. The absence of temporal structure in the hits reinforces the validity of the GARCH model for risk estimation, suggesting that it effectively captures the dynamics of market volatility without generating systematic bias.

Figure 25: Analysis of ACF for period 2010-2023 (HS)



This figure presents the ACF analysis of VaR hits obtained with the historical method (HS) for JPM, AIG and GS stocks. As observed, the autocorrelation values remain low and are mostly within the confidence bands, suggesting an absence of significant autocorrelation in the hits. This indicates that VaR violations are independent over time, thus respecting a key assumption of a reliable risk management model.

Table 15: Portmanteau tests (2009-2023)

Portmanteau tests	lb_stat (GARCH)	lb_pvalue (GARCH)	lb_stat (HS)	lb_stat (HS)
JPM	25.4733	0.0001	165.9301	0.0000
AIG	39.5501	1.840180e-07	186.2162	0.0000
GS	11.4421	0.0433	81.9771	3.330669e-16

Portmanteau test analysis reveals statistically significant results for all three investment banks, with extremely low p-values below the 5% threshold for both the GARCH and historical simulation (HS) models. For the GARCH model, JPM has a statistic of 25.47, AIG has 39.55, and GS has 11.44, with p-values of 0.00011, 0.00000018, and 0.0433, respectively. The results are even more pronounced with the HS approach, where the statistics are considerably higher (165.93 for JPM, 186.22 for AIG, and 81.98 for GS) with p-values of virtually zero. These results strongly suggest the presence of significant time dependence in VaR violations, indicating that both models may not fully capture the temporal dynamics of risks for these institutions.

Table 16: Portmanteau tests (2010-2023)

Portmanteau tests	lb_stat (GARCH)	lb_pvalue (GARCH)	lb_stat (HS)	lb_stat (HS)
JPM	25.4733	11.2868	165.9301	0.0000
AIG	40.14495	1.396153e-07	186.2162	0.0000
GS	17.7468	0.0033	81.9770	3.330669e-16

The Portmanteau (Ljung-Box) test presented in this table assesses the autocorrelation of VaR violations under the GARCH and historical (HS) models. For all three stocks (JPM, AIG, GS), the test statistics (lb_stat) are high, and the associated p-values are extremely low (close to zero), indicating a significant rejection of the null hypothesis of no autocorrelation. This suggests that VaR violations are not independent over time, which calls into question the assumption of well-calibrated models. The effect is particularly marked under the HS method, with much higher lb_stat values than under GARCH, indicating that the VaR estimated by the historical method has a stronger time dependence in the violations, which may limit its reliability for risk management.

Table 17: Box_pierce (2009-2023)

Box pierce	**bp_stat (GARCH)**	**bp_pvalue (GARCH)**	**bp_stat (HS)**	**bp_stat (HS)**
JPM	25.4478	1.141582e-04	165.7355	0.0000
AIG	39.5034	1.880481e-07	185.9933	0.0000
GS	11.4297	0.0435	81.8780	3.330669e-16

Analysis of the Box-Pierce test results over the period 2009-2023 confirms and strengthens the conclusions of the previous Portmanteau test. For the GARCH model, we observe significant test statistics for all three banks, with values of 25.45 for JPMorgan Chase, 39.50 for AIG, and 11.43 for Goldman Sachs, all associated with very low p-values (respectively 0.00011, 0.00000019, and 0.0435). The results are even more pronounced with the historical simulation (HS) approach, where the test statistics are significantly higher (165.74 for JPM, 186.00 for AIG, and 81.88 for GS) with p-values of practically zero. This consistency between the Box-Pierce and Portmanteau tests strengthens the conclusion that there is significant time dependence in VaR violations.

Table 18: Box_pierce (2010-2023)

Box pierce	**bp_stat (GARCH)**	**bp_pvalue (GARCH)**	**bp_stat (HS)**	**bp_stat (HS)**
JPM	21.4537	6.647826e-04	165.7355	0.0000
AIG	40.0975	1.427207e-07	185.9933	0.0000
GS	17.7229	0.0033	81.8780	3.330669e-16

The results show that for all three stocks (JPM, AIG, GS), the test statistics (bp_stat) are high and the associated p-values are extremely low, suggesting a significant rejection of the null hypothesis of no autocorrelation. As with the Portmanteau test, autocorrelation is more pronounced under the HS method, with particularly high test statistics and zero p-values, indicating a strong time dependence of VaR violations. This finding suggests that GARCH models provide a better capture of risk dynamics compared to the historical method, although adjustments may still be necessary to remove any residual dependence.

Table 19: LR backtest for period 2009-2023 (GARCH)

LR backtest	$Test_statistic_LR_IND	$Critical_Value_LR_IND	$Rejection_H0_LR_IND	$Pvalue_LR_IND
JPM	14.0642	3.8415	True	0.0002
AIG	10.0548	3.8415	True	0.0015
GS	14.0642	3.8415	True	0.0002

The results of the LR backtest for independence (IND) with the GARCH model over the period 2009-2023 show a categorical rejection of the null hypothesis of independence for all three financial institutions. Indeed, JPMorgan Chase and Goldman Sachs have an identical test statistic of 14.0642, while AIG displays a value of 10.0548, all significantly higher than the critical value of 3.8415. The very low p-values (0.0002 for JPM and GS, 0.0015 for AIG) confirm that the VaR violations are not independent over time. These results are consistent with the findings of the previous Portmanteau and Box-Pierce tests, suggesting a clustering of violations that could indicate an underestimation of risk during certain periods of stress or increased volatility.

Table 20: LR backtest for period 2009-2023 (HS)

LR backtest	$Test_statistic_LR_IND	$Critical_Value_LR_IND	$Rejection_H0_LR_IND	$Pvalue_LR_IND
JPM	37.6946	3.8415	True	8.273471e-10
AIG	44.5476	3.8415	True	2.482514e-11
GS	13.9757	3.8415	True	0.0002

The table presents the results of a likelihood test (LR backtest) over the period 2009-2023 for three financial institutions: JPM, AIG and GS. The null hypothesis (H0) states that the risk model used is well calibrated, i.e. that the risk predictions are consistent with actual observations. The test statistics obtained (37.6946, 44.5476 and 13.9757) largely exceed the critical value of 3.8415, leading to the rejection of H0 for all institutions. The extremely low p-values (< 0.05) confirm this conclusion. This suggests that the risk models used by JPM, AIG and GS are not well calibrated and could underestimate or overestimate financial risks over the period studied.

Table 21: LR backtest for period 2010-2023 (GARCH)

LR backtest	$Test_statistic_LR_IND	$Critical_Value_LR_IND	$Rejection_H0_LR_IND	$Pvalue_LR_IND
JPM	13.3926	3.8415	True	0.0003
AIG	17.7547	3.8415	True	2.513023e-05
GS	13.3926	3.8415	True	0.0003

The results show that for all stocks (JPM, AIG, GS), the test statistic exceeds the critical value of 3.8415, and the p-values are very low (less than 0.05). Therefore, the null hypothesis of independence of violations is rejected for all stocks. This indicates that VaR violations exhibit time dependence under the GARCH model, suggesting that the latter does not fully capture the risk dynamics and may require improvements, such as a more sophisticated model or consideration of long memory.

Table 22: LR backtest for period 2010-2023 (HS)

LR backtest	$Test_statistic_LR_IND	$Critical_Value_LR_IND	$Rejection_H0_LR_IND	$Pvalue_LR_IND
JPM	37.6946	3.8415	True	8.273471e-10
AIG	44.5476	3.8415	True	2.482514e-11
GS	13.9757	3.8415	True	0.0002

The results of the likelihood test for conditional independence (LR_IND) under the Historical (HS) method indicate that for all stocks (JPM, AIG, GS), the test statistic significantly exceeds the critical value of 3.8415. The p-values are extremely low (close to zero), leading to the rejection of the null hypothesis of independence of VaR violations. This means that VaR violations under the HS method exhibit strong autocorrelation, indicating a marked time dependence.

7.3. Testing the CC hypothesis

Table 23: LR backtest for CC htpothesis for period 2009-2023 (GARCH)

LR backtest	$Test_statistic_LR_CC	$Critical_Value_LR_CC	$Rejection_H0_LR_CC	$Pvalue_LR_CC
JPM	16.3044	5.9915	True	5.393833e-05
AIG	13.0455	5.9915	True	0.0003
GS	5.9183	5.9915	False	0.0149

The table presents the results of the likelihood test (LR backtest) for the conditional coverage (CC) hypothesis over the period 2009-2023 using a GARCH model. The null hypothesis H0 assumes that the risk model correctly captures the conditional dynamics of VaR violations.

For JPM and AIG, the test statistics (16.3044 and 13.0455) are well above the critical value of 5.9915, leading to the rejection of H0 with very low p-values. This suggests that the GARCH model fails to correctly capture the conditional risk dynamics for these institutions.

In contrast, for GS, the test statistic (5.9183) is slightly below the critical value, and the p-value (0.0149) is relatively higher. Thus, H0 is not rejected for GS, indicating that the GARCH model could be appropriate for this institution in terms of conditional coverage.

Table 24: LR backtest for CC htpothesis for period 2009-2023 (HS)

LR backtest	$Test_statistic _LR_CC	$Critical_Value_LR _CC	$Rejection_H0 _LR_CC	$Pvalue_LR_CC
JPM	37.6946	5.9915	True	8.273471e-10
AIG	44.6239	5.9915	True	2.387657e-11
GS	14.1587	5.9915	True	0.0002

The table presents the results of the likelihood test (LR backtest) for the conditional coverage (CC) hypothesis over the period 2009-2023 using the historical method (HS). The null hypothesis H0 states that the risk model correctly captures the time dependence of VaR violations.

The test statistics for JPM (37.6946), AIG (44.6239) and GS (14.1587) greatly exceed the critical value of 5.9915, leading to the rejection of H0 for all institutions. The extremely low p-values (< 0.05) confirm this conclusion.

These results suggest that the historical simulation-based model fails to correctly model the conditional risk dynamics for JPM, AIG and GS. It may therefore be necessary to improve or revise the approach used to better capture the structure of VaR violations.

Table 25: LR backtest for CC htpothesis for period 2010-2023 (GARCH)

LR backtest	$Test_statistic _LR_CC	$Critical_Valu e_LR_CC	$Rejection_H0 _LR_CC	$Pvalue_LR_CC
JPM	15.4179	5.9915	True	8.616904e-05
AIG	25.6352	5.9915	True	4.124454e-07
GS	2.2259	5.9915	False	0.1357

The results of the likelihood test for the joint condition (LR_CC) under the GARCH model show notable differences between the assets. For JPM and AIG, the test statistic is well above the critical value of 5.9915, with very low p-values, leading to the rejection of the null hypothesis of model adequacy. This means that the GARCH model does not fully capture the risk dynamics for these assets. In contrast, for GS, the test statistic is below the critical value and the p-value is relatively high (0.1357), suggesting that the null hypothesis cannot be rejected. In other words, the model seems to perform better for GS than for JPM and AIG.

Table 26: LR backtest for CC htpothesis for period 2010-2023 (HS)

LR backtest	$Test_statistic_LR_CC	$Critical_Value_LR_CC	$Rejection_H0_LR_CC	$Pvalue_LR_CC
JPM	37.6946	5.9915	True	8.273471e-10
AIG	44.6239	5.9915	True	2.387657e-11
GS	14.1587	5.9915	True	0.0002

The results of the likelihood test for the joint condition (LR_CC) under the historical method (HS) indicate a systematic rejection of the null hypothesis for all three assets (JPM, AIG and GS). The test statistics are significantly higher than the critical value of 5.9915, and the p-values are extremely low, confirming the inadequacy of the HS model to adequately capture the risk. Unlike the results observed with the GARCH model, where GS seemed to fit better, here even GS fails the test, suggesting a generalized underperformance of the HS model for the period 2010-2023.

8. CONCLUSION

In this study, we analyzed the market risks of three major financial institutions – JPMorgan Chase, American International Group (AIG) and Goldman Sachs – over the period 2007-2023. To do this, we used two complementary approaches: Historical Simulation (HS) and GARCH (1,1) models, to estimate and anticipate key indicators such as Value-at-Risk (VaR) and Expected Shortfall (ES).

Our results show that GARCH models are more responsive to market shocks and provide better consideration of extreme risks, unlike HS models which, although intuitive, rely only on past data and struggle to adapt to sudden market variations. Backtesting tests confirm that HS models remain relevant in terms of overall coverage but have weaknesses when it comes to assessing the dependence of VaR exceedances and their conditional nature.

Time series analysis highlights several features specific to financial markets, such as volatility persistence, skewness of distributions, and the presence of fat tails. This reinforces the idea that models such as GARCH, which dynamically adjust volatility, are essential for effective risk management. Furthermore, statistical tests of autocorrelation (Portmanteau and Box-Pierce) have revealed dependencies in VaR exceedances, suggesting that some models do not fully capture risk dynamics. Furthermore, likelihood ratio (LR) tests have rejected the independence assumption, highlighting some structural limitations of the assumptions underlying VaR. The financial crises of 2008 and the pandemic of 2020 have highlighted the importance of using models that can adapt quickly to extreme market conditions. In this sense, coverage condition (CC) tests show that the

HS method tends to overestimate risks compared to the GARCH model, although the latter is not free of limitations.

In conclusion, our results suggest that GARCH models are better suited to anticipate and manage risks in periods of high volatility. However, a combined HS and GARCH approach could offer a more balanced assessment, combining the simplicity of nonparametric methods with the responsiveness of parametric models. In the future, it would be relevant to explore even more advanced approaches, such as stochastic volatility models or artificial intelligence, to improve the accuracy and robustness of risk forecasts.

REFERENCES

Alexander, C. (2008). *Market risk analysis, volume II: Practical financial econometrics*. John Wiley & Sons.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). *Coherent measures of risk*. *Mathematical Finance, 9*(3), 203–228. https://doi.org/10.1111/1467-9965.00068

Bollerslev, T. (1986). *Generalized autoregressive conditional heteroskedasticity*. *Journal of Econometrics, 31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Brownlees, C., & Engle, R. (2016). *SRISK: A conditional capital shortfall measure of systemic risk*. *The Review of Financial Studies, 30*(1), 48–79. https://doi.org/10.1093/rfs/hhw060

Christoffersen, P. (1998). *Evaluating interval forecasts*. *International Economic Review, 39*(4), 841–862. https://doi.org/10.2307/2527341

Engle, R. F. (1982). *Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation*. *Econometrica, 50*(4), 987–1007. https://doi.org/10.2307/1912773

Fruzzetti, M., Nasti, D., & Orlandi, M. (n.d.). *The backtesting of value-at-risk and expected shortfall: Evidence from the Bank of Italy's foreign reserves portfolio*. *Evolving Practices in Public Investment Management*. Bank of Italy.

Hué, S. (n.d.). *Lessons in financial econometrics*. Aix-Marseille School of Economics.

Hull, J. C. (2018). *Options, futures, and other derivatives* (10th ed.). Pearson.

Jorion, P. (2007). *Value at risk: The new benchmark for managing financial risk* (3rd ed.). McGraw-Hill.

J.P. Morgan. (1996). *RiskMetrics—Technical document*. Morgan Guaranty Trust Company.

Kupiec, P. H. (1995). *Techniques for verifying the accuracy of risk measurement models*. *The Journal of Derivatives, 3*(2), 73–84. https://doi.org/10.3905/jod.1995.407942

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools* (2nd ed.). Princeton University Press.

New Trading. (n.d.). *Backtester une stratégie de trading: Guide complet*. Retrieved from https://www.newtrading.fr/backtester-strategie-trading

Pritsker, M. (2006). *The hidden dangers of historical simulation*. *Journal of Banking & Finance, 30*(2), 561–582. https://doi.org/10.1016/j.jbankfin.2005.04.015

Yahoo Finance. (n.d.). *JPMorgan Chase & Co. (JPM) historical data*. Retrieved from https://finance.yahoo.com/quote/JPM/history

Yahoo Finance. (n.d.). *American International Group, Inc. (AIG) historical data*. Retrieved from https://finance.yahoo.com/quote/AIG/history

Yahoo Finance. (n.d.). *Goldman Sachs Group, Inc. (GS) historical data*. Retrieved from https://finance.yahoo.com/quote/GS/history

APPENDIX

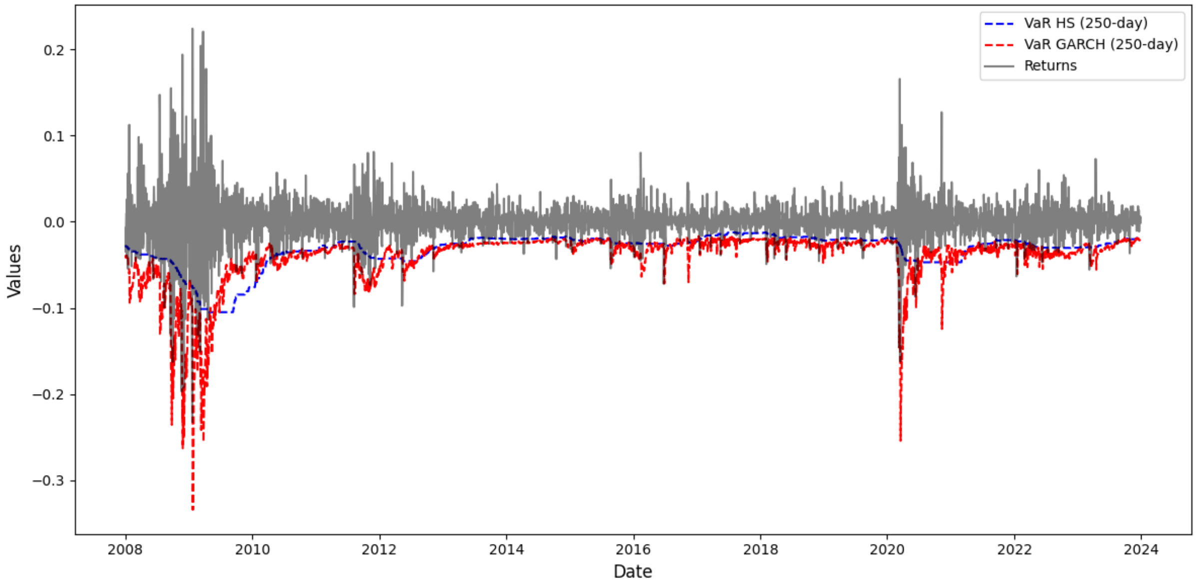


Fig. 1 — VaR Forecasts for JPM using a 250-day Rolling Window

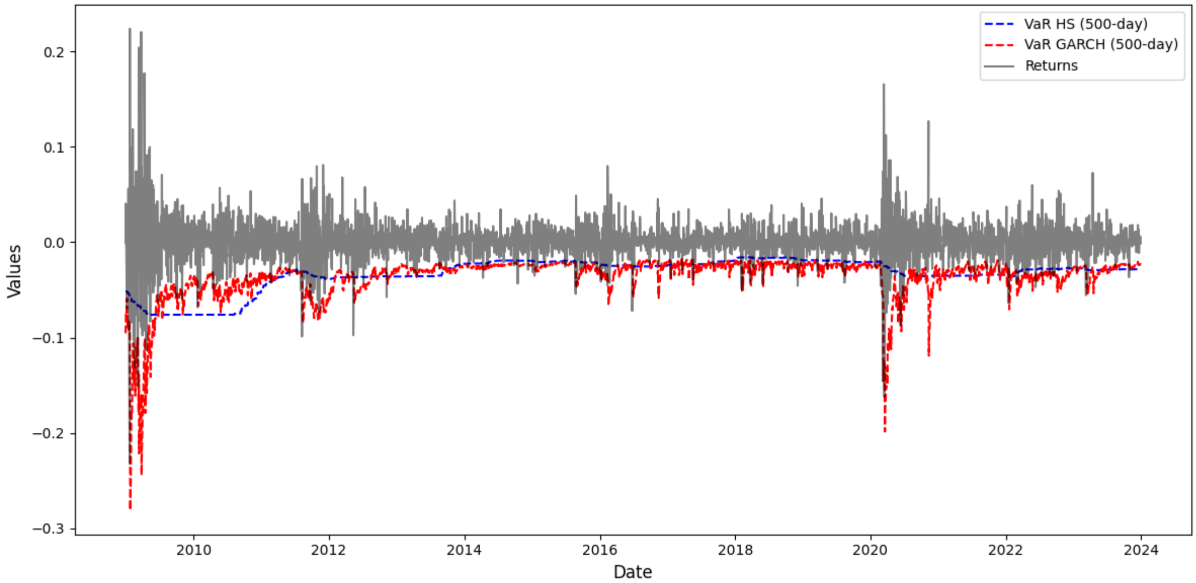


Fig. 2 — VaR Forecasts for JPM using a 500-day Rolling Window

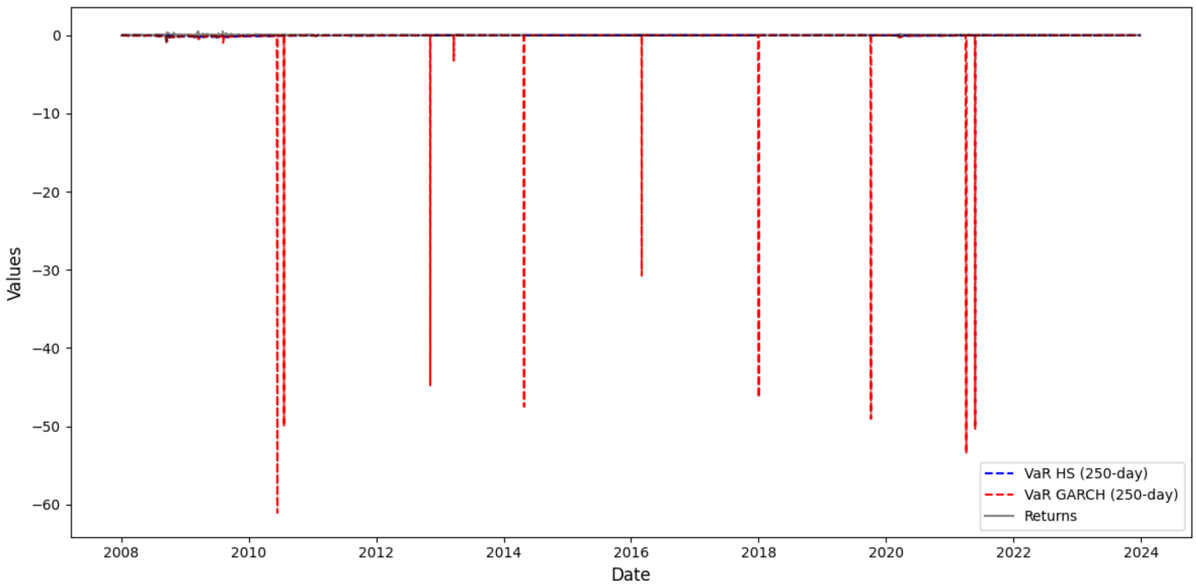


Fig. 3 — VaR Forecasts for AIG using a 250-day Rolling Window

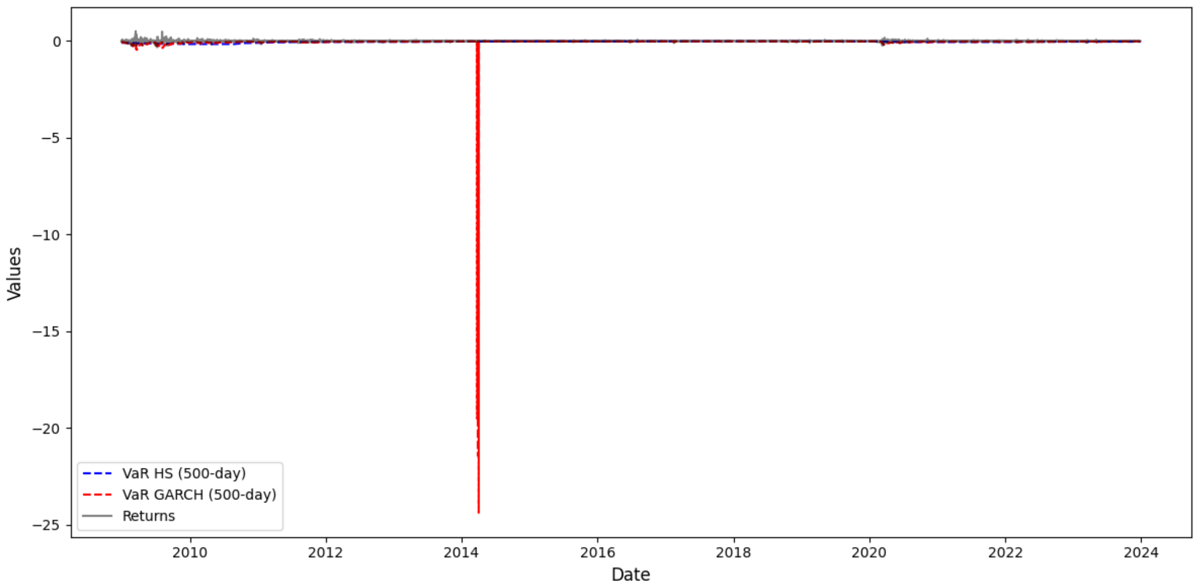


Fig. 4 — VaR Forecasts for AIG using a 500-day Rolling Window

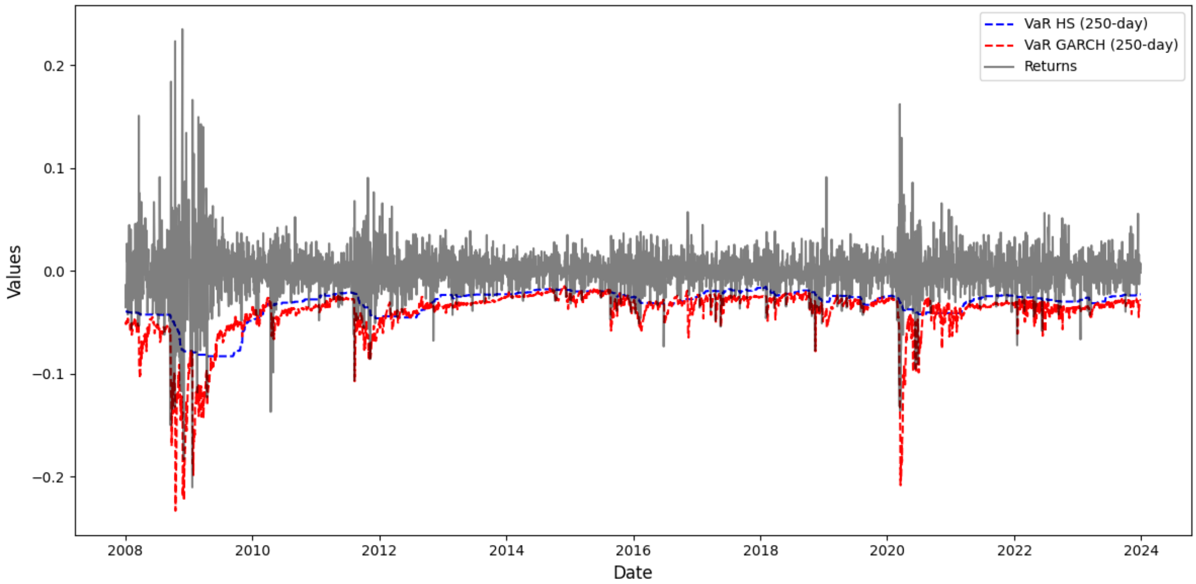


Fig. 5 — VaR Forecasts for GS using a 250-day Rolling Window

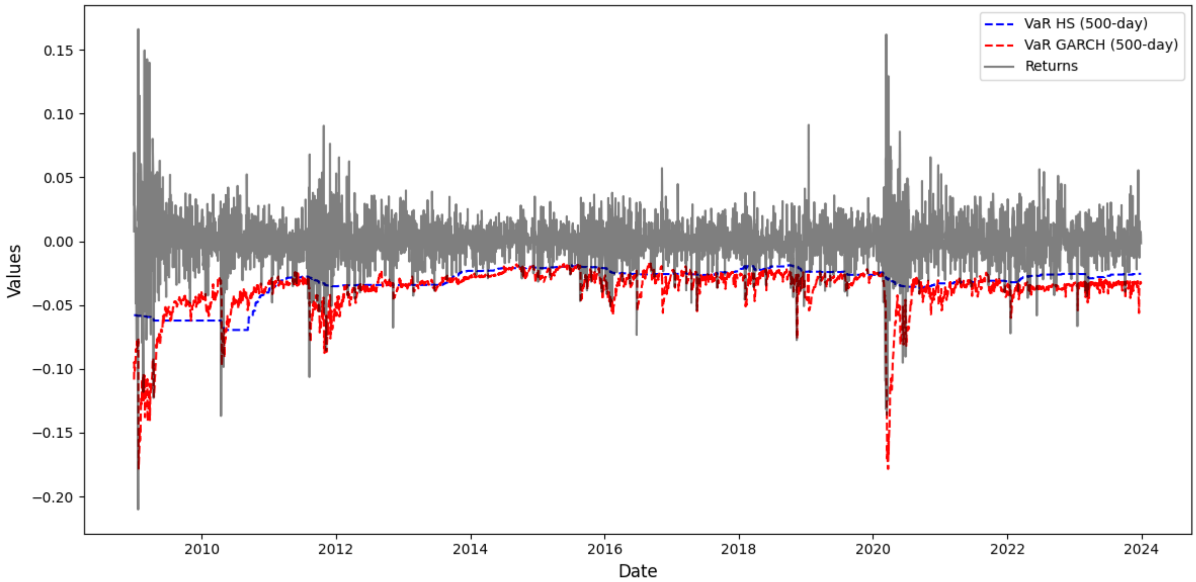


Fig. 6 — VaR Forecasts for GS using a 500-day Rolling Window

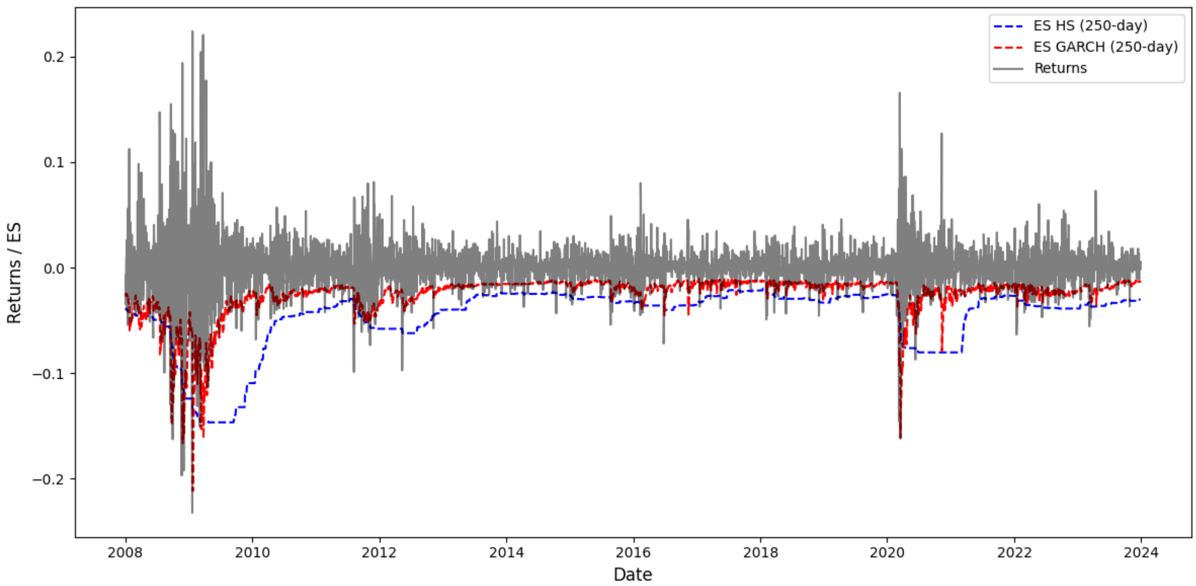


Fig. 7 — Forecast for ES using a 250-day Rolling Window for JPM

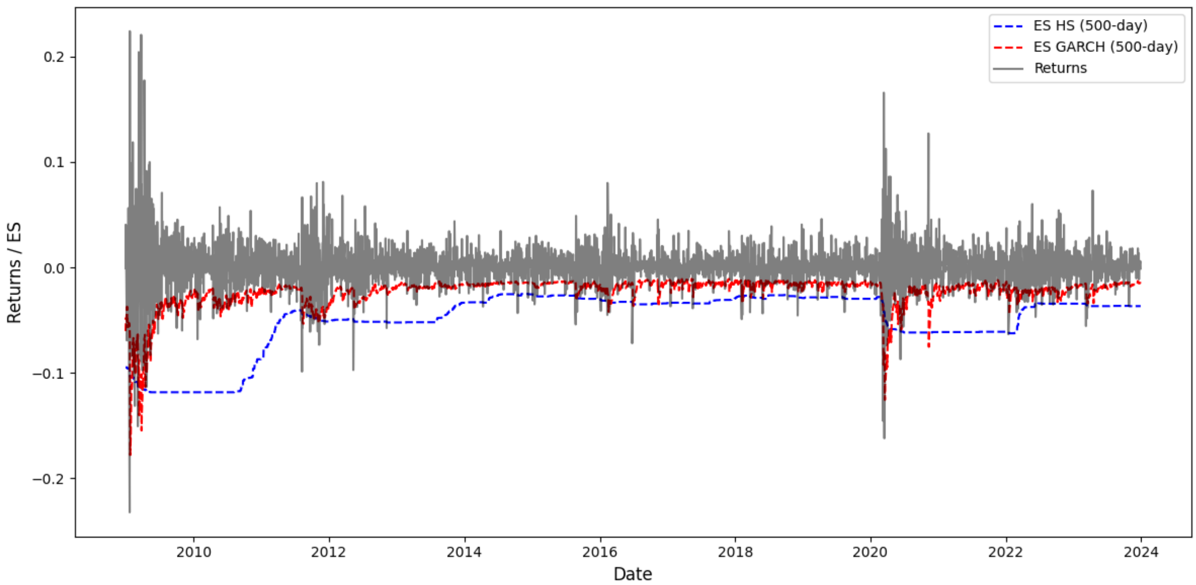


Fig. 8 — Forecast for ES using a 500-day Rolling Window for JPM

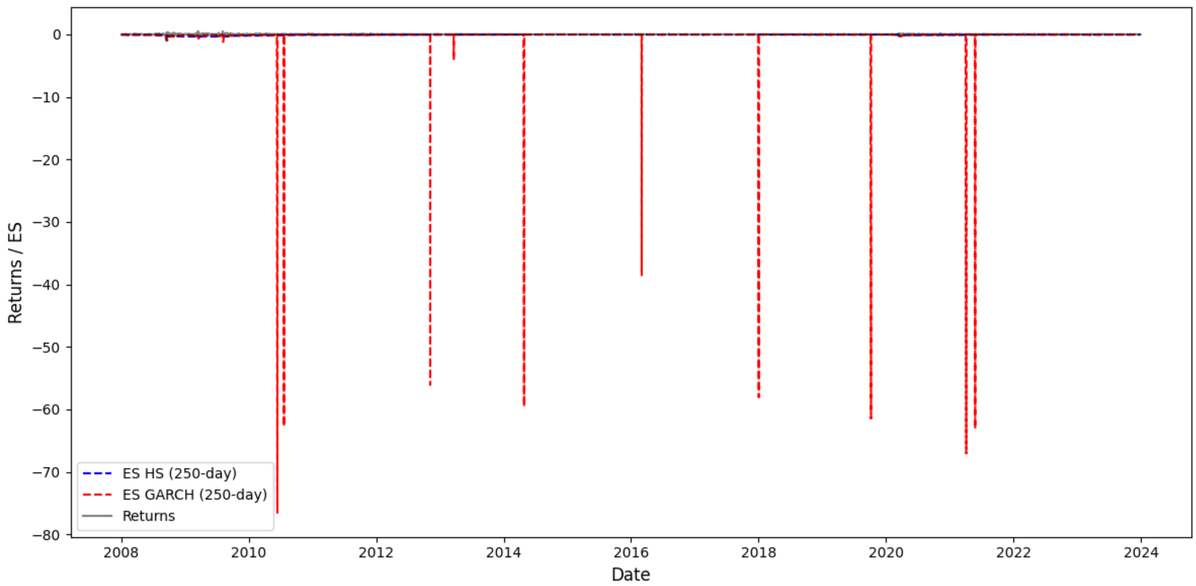


Fig. 9 — Forecast for ES using a 250-day Rolling Window for AIG

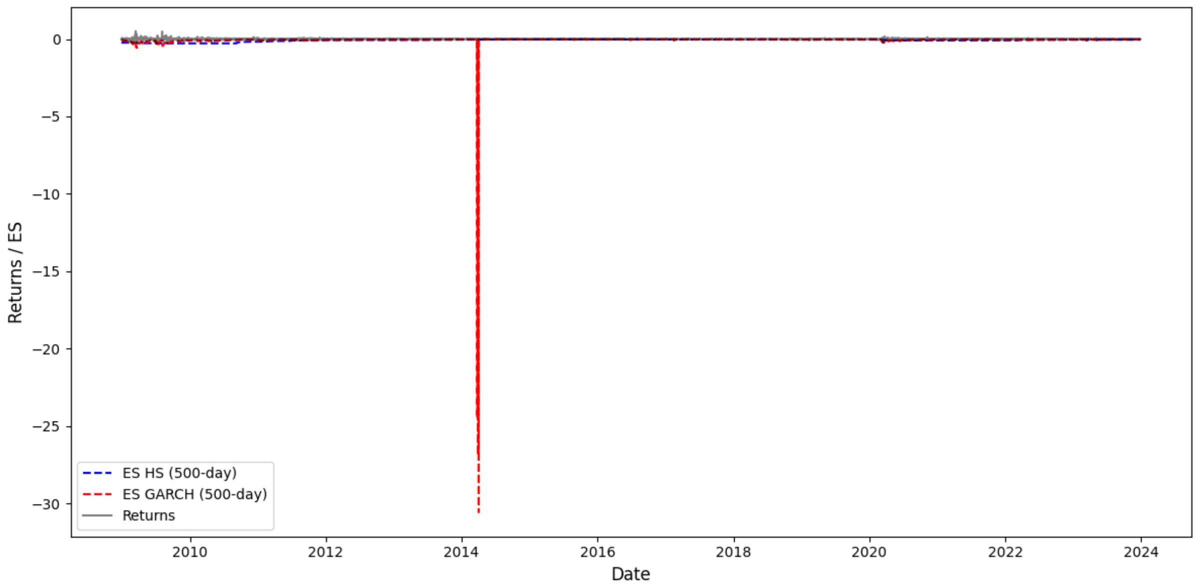


Fig. 10 — Forecast for ES using a 500-day Rolling Window for AIG

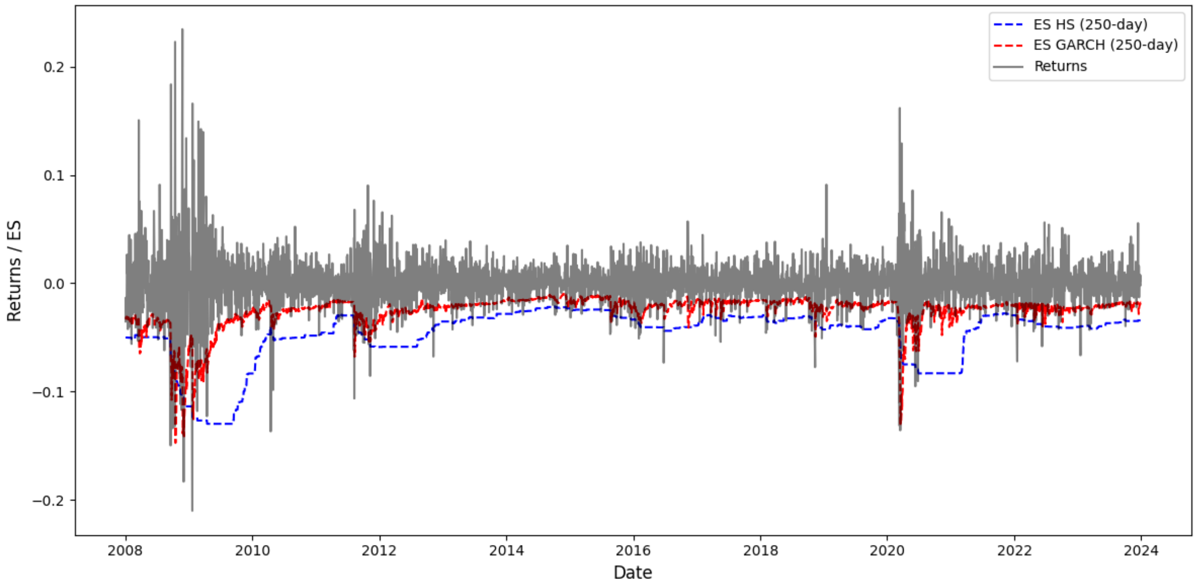


Fig. 11 — Forecast for ES using a 250-day Rolling Window for GS

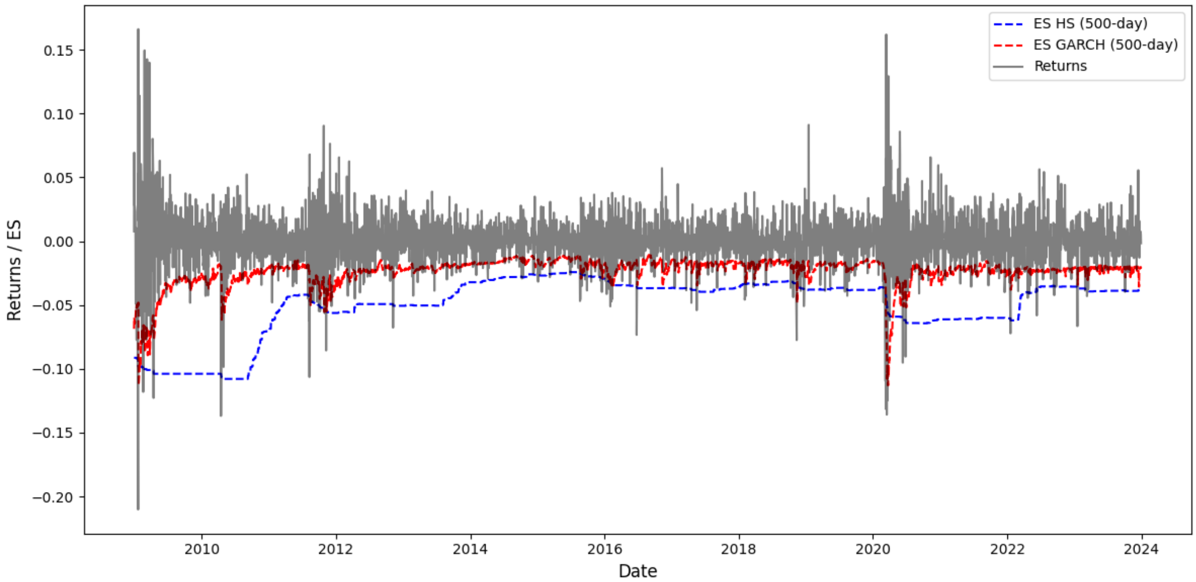


Fig. 12 — Forecast for ES using a 500-day Rolling Window for GS