

Forecasting Value-At-Risk Using Different Volatility Models

MSc Data-Intensive Analysis, University of St Andrews

Cristobal Valdivia*

2023-08-15

Abstract

Forecasting Value-at-Risk (VaR) accurately is pivotal for prudent financial risk management, especially when considering the fluctuations of major stock indices like the S&P500. This research addresses the challenges inherent in commonly-used simplistic VaR models, such as those that assume constant volatility and normality in returns distribution. Drawing upon five years of S&P500 closing price data sourced from Yahoo Finance, this study embarks on refining the VaR estimation process. The data, analyzed using a 252-day window, undergoes a series of transformations through advanced volatility models that consider the non-normal distribution of returns, characterized by pronounced fat tails. The journey commences with the constant volatility model, explores the leverage effect, and evolves into more sophisticated GARCH-based models. Furthermore, a dynamic model selection technique is introduced to amalgamate the strengths of different models. Back-tests with over 1,006 VaR predictions confirm the superior efficacy of the latter models, underscoring the importance of understanding the intricate dynamics of financial returns. In essence, this dissertation offers a comprehensive progression from rudimentary to refined VaR models, promoting enhanced accuracy in forecasting financial risk.

*Supervised by: Valentin Popov

Contents

Introduction	3
1. First: The Data!	3
2. Fundamentals of Value-At-Risk: The simplest volatility model (constant)	3
3. A brief empirical study: Returns are normally distributed?	5
4. Understanding the dynamics of financial returns: non-constant volatility models	8
Weighting Schemes	8
Model 1: Exponentially weighted moving average (EWMA)	8
Model 2: Generalized Auto-regressive Conditional Heteroskedasticity (GARCH(1,1))	12
5. Leverage Effect: Value-at-Risk by Addressing Asymmetric Distributions	16
The Leverage Effect	17
Model 3: GJR-GARCH(1,1)	18
Model 4: Exponential GARCH Model, EGARCH(1,1)	22
6. Boosting leverage effect volatility models	25
Final adjustments GJR-GARCH(1,1): Optimum t-Student distribution (grid-search approach)	25
Model 5: The best of both worlds (GJR-GARCH and EGARCH)	28
7. Back-Testing: Assess models performance	30
Conclusion	33
References	34

Introduction

In contemporary finance, the ability to forecast risk is a linchpin for the stability and success of financial institutions, especially in the context of economic cyclicality. At the heart of this predictive endeavour lies the Value-At-Risk (VaR) approach, a pivotal risk metric that encapsulates the maximum potential loss an investment portfolio could face over a specific period for a given confidence interval. Its significance is underscored by its capacity to condense complex risk metrics, such as Greek letters, into a singular, interpretable number. However, the integrity of VaR forecasts hinges on the robustness of the underlying volatility models, with the recognition that simplistic assumptions can prove calamitous in the real-world application. This research, therefore, seeks to amplify the precision of VaR estimations for the S&P500 Index daily returns by transitioning from rudimentary models, like constant volatility, to more sophisticated GARCH-based processes.

The S&P500, a benchmark index of the American stock market, offers rich insights into the patterns and irregularities of financial returns. By leveraging five years' worth of S&P500 close price data, sourced from Yahoo Finance, this dissertation embarks on an analytical journey to conceive a robust VaR model, tailored to predict the 5th percentile of the loss distribution. An inherent challenge, however, is the non-normal distribution of financial returns, often characterized by fat tails, necessitating the adoption of more complex volatility estimation techniques.

Methodologically, this work is rooted in data-intensive analysis. Utilizing a 252-day window size for model fitting, the subsequent model validations encompass a hefty 1,006 VaR predictions, juxtaposed against the actual S&P500 index returns. Through this iterative process, the research uncovers the nuances of financial return dynamics, exploring the relevance of the leverage effect, and eventually fusing multiple volatility models through dynamic model selection techniques for optimal outcomes.

The subsequent chapters unravel this analytical tapestry: commencing with an exploration of the data, segueing into foundational discussions around VaR, and progressively delving into the intricacies of volatility modeling, culminating in exhaustive back-tests to ascertain the efficacy of the proposed models.

In essence, this dissertation elucidates the journey from basic VaR estimations, predicated on rudimentary volatility assumptions, to intricate models that echo the multifaceted dynamics of financial returns, fostering a richer, more accurate risk forecasting paradigm.

1. First: The Data!

Financial data used was collected using *quantmod* in R, which is a popular library for quantitative financial modeling analysis. This library provides functions for downloading, manipulating, and analyzing financial data, from various sources and in our work we will use Yahoo Finance as the main source. For the further analysis, we download a time series from 11th June 2018 to 11th June 2023 (5 years, 1259 observations) for the Standard & Poor's 500 Index which is a stock market index tracking the stock performance of 500 of the largest companies listed on stock exchanges in the United States.

Main idea of this work is manipulate this Data by using programming languages, such as R and python, to build some mathematics and statistics models to estimate a daily Value-At-Risk for S&P500 Index. We begin with the most simple volatility models (constant) and as the work evolve, increase the complexity of those with the aim to improve accuracy of VaR.

2. Fundamentals of Value-At-Risk: The simplest volatility model (constant)

As the first step, starts from the simplest VaR model which has two main assumptions:

- (1) A daily change in S&P500 Index prices is normally distributed (returns are normally distributed)
- (2) Standard deviation of returns will be constant (constant volatility)

When the loss, measured as a daily percentage change, has a mean of μ and a standard deviation of σ , VaR is defined as:

$$\mathbf{VaR}_{(\alpha, \Delta t)} = \mu + \sigma \cdot Z_{\alpha} \quad (1)$$

where:

α is the confidence level, or percentile

Δt is the time horizon

Z is the inverse cumulative Normal distribution

We must choose two parameters to compute the VaR for the S&P500 Index; (a) confidence level and (b) time horizon.

All models in this work will use a confidence level of 95% and as S&P 500 Index are very liquid choose a short time horizon (1 day). After selected those parameters for VaR formula shown in equation (1), VaR depends on the mean of daily returns (μ) and the volatility (σ) and we can rewrite equation (1) as:

$$\mathbf{VaR}_{(0.95, 1_{day})} = \mu + \sigma \cdot Z_{0.95} \quad (2)$$

Using the full data (5 years), the empirical average daily return for the S&P500 Index (μ) was 0.04%. As Δt is small (1 day), mean of daily returns is often assumed to be zero ¹, and therefore equation (2) simplified and VaR will depends only in the expected 1 day volatility as shows the next formula:

$$\mathbf{VaR}_{(0.95, 1_{day})} = \sigma \cdot Z_{0.95} \quad (3)$$

In our first step, we assumed a constant volatility compute as simple as the standard deviation of the S&P500 daily returns showed in the first 5 years of the Data. Constant volatility assume equal weights for all returns observed in the dataset as shown in equation (4)

$$\sigma_n = \sqrt{\frac{1}{m} \sum_{i=1}^m (u_{n-i} - \bar{u})^2} \quad (4)$$

where:

n is the time horizon (1 day)

m is the number of observations

u_{n-i} is the return observed in S&P 500 Index i days ago

\bar{u} is the average daily return, which on this work will assume to be 0

By using all our dataset, we have all those values

$$\sigma = \sqrt{\frac{1}{1259} \sum_{i=1}^{1259} (u_{1-i})^2} = 1.37\%$$

And finally using equation (3) we can forecast a 95% confidence level VaR for the next day (12th June 2023) assuming returns follows a normal distribution with a mean of 0 and a standard deviation of 1.37% (daily

¹Hull, J.C., 2018. 'Choice of Parameters for VaR and ES'. In: Risk Management and Financial Institutions. 5th ed. Hoboken, New Jersey: John Wiley & Sons, Inc., pp.278.

average volatility)

$$\begin{aligned}
 \text{VaR}_{(0.95,1_{day})} &= \sigma \cdot Z_{0.95} \\
 &= 1.37\% \cdot 1.64 \\
 &= 2.25\%
 \end{aligned}$$

Using this simple VaR model we are 95% confident that we will not lose more than 2.25% in 1 day forward if we invest on the S&P 500 Index. Figure 1 shows the normal distribution assumes to model daily S&P500 Index daily returns and the VaR computed before.

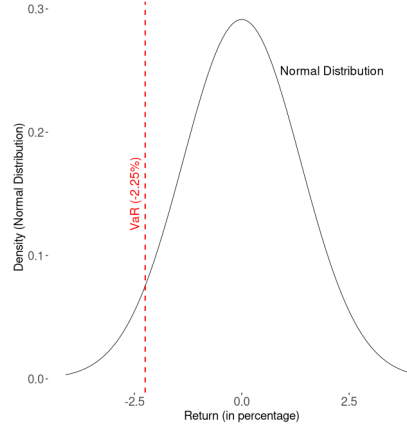


Figure 1: Value-At-risk with a confidence level of 95% using a Normal distribution to model S&P500 daily returns

Returns are normally distributed is realistic? The aim of the next section is to answer this particular question by contrast the empirical returns distributions of the S&P 500 Index with a normal distribution showed in Figure 1

3. A brief empirical study: Returns are normally distributed?

At the beginning, JPMorgan and RiskMetrics Group proposed a method for calculating the potential downside risk of a particular investment due to market fluctuations (1994)². This was the first time that a Value-At-Risk model was launched to the public. Since its introduction, VaR model has undergone refinements and enhancements but the basics in VaR model assumes that returns of any financial security follows a normal distribution. On this section we will study the empirical distribution of returns for the S&P 500 Index and contrast this with the normal distribution proposed before $\mathcal{N}(\mu = 0, \sigma = 1.37\%)$ (Figure 1)

Stock prices, fluctuate due various factors such as market conditions, company performance, investor sentiment, and economic events. In order to understand these fluctuations first analyze the S&P500 Index prices from June 2018 until June 2023 (5 years) and the daily returns showed on those days.

First as illustrates Figure 2 on the left hand side (price evolution in dollars) the S&P500 Index has a positive long-term trend, but in the short-term the price showed high fluctuations which are more clearly in the right hand chart (daily returns in percentage). From the first months of 2020, due to COVID-19 crisis, price declined drastically and lose all the gains from the last two years and this produce a significantly spike on returns in absolute values which was traduced in a higher volatility environment for the S&P500 Index.

We suspects that a normal distribution cannot explain well these short-term fluctuations. Unexpected significant fluctuations on the price generally are *outliers* events and by using a normal distribution those movements on the price could be underestimate, and therefore the VaR forecast.

²J.P. Morgan/Reuters Financial, 1996. RiskMetrics - Technical Document. 4th ed. [pdf] New York: J.P. Morgan. Available at: Link [Accessed June 14th, 2023]

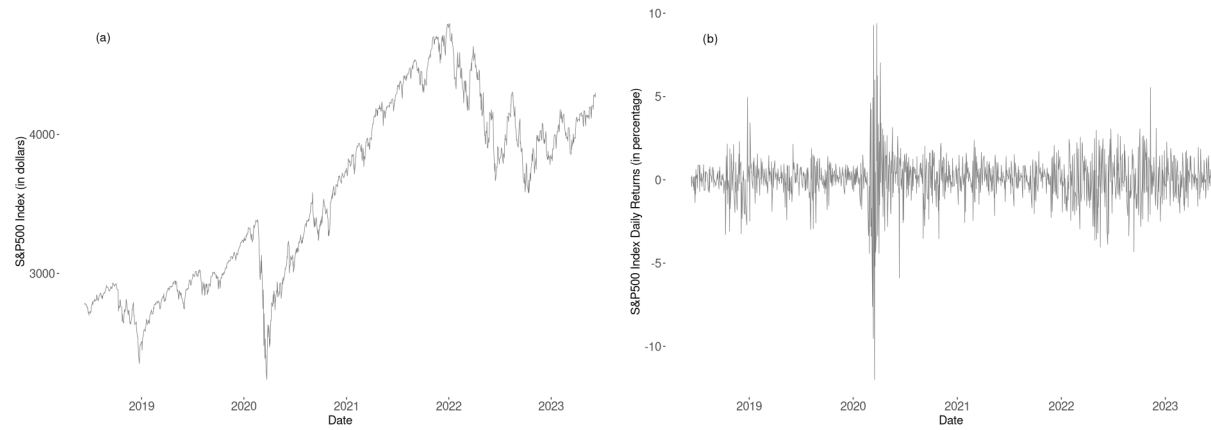


Figure 2: (a) Close Price in dollars and (b) daily returns in percentage

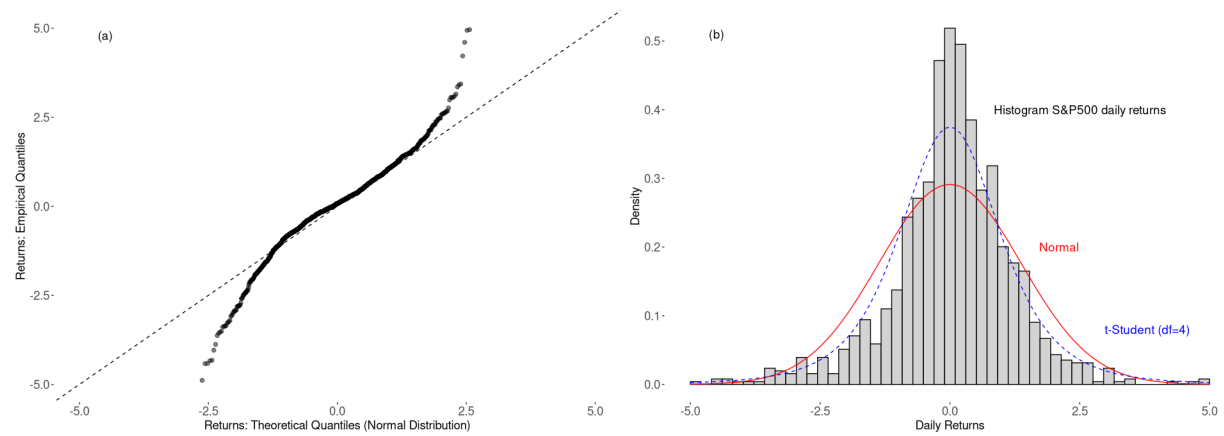


Figure 3: (a) Theoretical quantiles for a Normal Distribution vs Empirical quantiles daily returns and (b) Histogram of daily returns vs Normal distribution and t-Student distribution.

With the aim to understand the effectiveness fit of a normal distribution on the S&P500 Index daily returns, Figure 3 showed a quantile-quantile plot (QQ-plot) on the left hand side and on the right a S&P500 Index histogram daily returns with two statistical distributions; (a) normal distribution and (b) t-Student. A QQ-plot compares the quantiles of the observed returns (our sample for the S&P500 Index) with the quantiles of a theoretical distribution ($\mathcal{N}(\mu = 0, \sigma = 1.37\%)$) to assess their similarities. If a normal distribution quantiles are equals to the sample quantiles a QQ-plot should show all observations from the sample (black dots) aligned with the black dotted line. As we expected before, on the extremes values (tails) normal distribution underestimated those events and empirical returns has wider tails, therefore using a normal distributions implied an underestimation for the VaR.

Additionally exist a test statistic called Shapiro-Wilk used to determine whether a dataset, in our case S&P500 Index returns, follows a normal distribution. A null hypothesis for this test (H_0) is that S&P500 returns are normally distributed. Using R software, the test-statistic (W) was 0.88 and the p-value associated was less than $2.2e-16$. As the p-value was less than 5%, we reject the null and accept the alternative with a 95% confidence level and concludes that our S&P500 Index daily returns are not normally distributed. *But, how fix this issue?* By inspection the right hand side chart on Figure 3, fortunately exist other distribution with a better fit on the tails than a normal distribution which is a t-Student with 4 degrees of freedom.

By replace $Z_{0.95}$ with $t_{(0.95, df=4)}$ on equation (3):

$$\text{VaR}_{(0.95, 1_{day})} = \sigma \cdot t_{(0.95, df=4)} \quad (5)$$

$$\begin{aligned} \text{VaR}_{(0.95, 1_{day})} &= \sigma \cdot t_{(0.95, df=4)} \\ &= 1.37\% \cdot 2.13 \\ &= 2.92\% \end{aligned}$$

As we expected using a t-Student instead of a normal distribution the VaR increased from 2.25% to 2.92% (+67 basis points) as shown in Figure 4

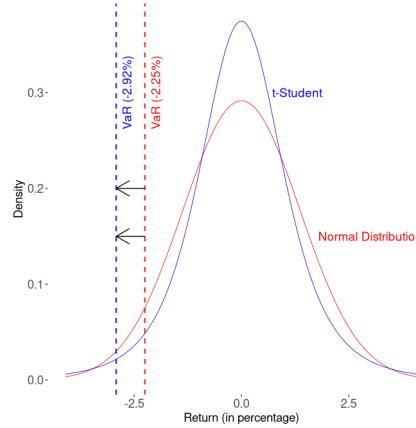


Figure 4: S&P500 Index daily returns vs Value-At-Risk using a Normal distribution and a t-Student with 4 degrees of freedom.

Assumes that returns behavior has a constant volatility could be a problem in our VaR model. When daily volatility of S&P500 spikes the probability of underestimate the VaR using a constant sigma could be significantly high. On the other side when volatility remains stable and in low levels we could overestimate the VaR using this simple approach.

If we compare graphically the empirical returns evolution of the S&P500 Index during this 5 years of data with a simulated iid data from a Normal distribution and t-Student where the parameters have been determined by fitting the models to the S&P500 daily returns (Figure 5), we can conclude that empirical volatility of

daily returns contains clusters which means that volatility appears to vary over time and therefore assuming constant volatility is not a reliable assumption as showed the simulated iid daily returns in Figure 5 (b)

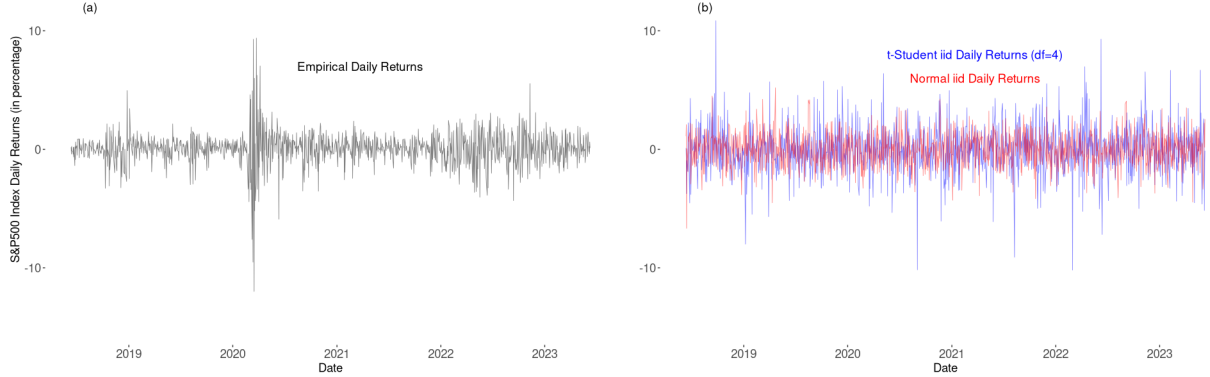


Figure 5: (a) S&P500 Index empirical daily returns and (b) Normal and t-Student iid daily returns simulated

Focus for the next sections of this report is to build some robust volatility models in order to incorporate the dynamics of empirical returns (clusters) with the aim to improve the quality of our Value-At-Risk models.

4. Understanding the dynamics of financial returns: non-constant volatility models

First start with the most basics non-constant variance models, such as exponentially weighted moving average (EWMA) and Generalized Auto-regressive Conditional Heteroskedasticity (GARCH).

Weighting Schemes

As we observed before, equation (4) gives equal weights to each daily return experienced by the S&P500 Index in our data (u_{n-i}). Our objective is to estimate the volatility of the S&P500 one day ahead to compute the VaR. Therefore makes sense to give more weight to recent data (returns) as shows the following equation:

$$\sigma_n^2 = \sum_{i=1}^m \alpha_i u_{n-i}^2 \quad (6)$$

where:

α_i refers to the weight given to the return i days ago.

and weights must sum to one:

$$\sum_{i=1}^m \alpha_i = 1$$

On the next three models we will use the main idea of the weighting schemes, with the aim to give less weight to older returns of the S&P500 Index and estimate volatility (σ)

Model 1: Exponentially weighted moving average (EWMA) ³

An EWMA model is a particular weighting scheme proposed in equation (6) The weights, as the name suggest, decreased exponentially as move back through time. Specifically $\alpha_{i+1} = \lambda \alpha_i$, where λ is a constant between 0 and 1.

³Hull, J.C., 2018. 'The Exponentially Weighted Moving Average Model'. In: Risk Management and Financial Institutions. 5th ed. Hoboken, New Jersey: John Wiley & Sons, Inc., pp.225-228.

In general terms to updating volatility estimates using a EWMA model, the formula is:

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2 \quad (7)$$

where:

σ_n^2 is the squared volatility estimate at the end of day $n - 1$ for day n

σ_{n-1}^2 is the squared volatility estimate made at the end of day $n - 2$ for day $n - 1$

u_{n-1}^2 is the squared most recent daily return observed

By substitute σ_{n-1}^2 on equation (7):

$$\begin{aligned} \sigma_n^2 &= \lambda(\lambda \sigma_{n-2}^2 + (1 - \lambda) u_{n-2}^2) + (1 - \lambda) u_{n-1}^2 \\ &= \lambda^2 \sigma_{n-2}^2 + (1 - \lambda)(u_{n-1}^2 + \lambda u_{n-2}^2) \end{aligned}$$

And similar way for σ_{n-2}^2 gives:

$$\sigma_n^2 = \lambda^3 \sigma_{n-3}^2 + (1 - \lambda)(u_{n-1}^2 + \lambda u_{n-2}^2 + \lambda^2 u_{n-3}^2)$$

Continuing in this way, we can write equation (7) as:

$$\sigma_n^2 = (1 - \lambda) \sum_{i=1}^m \lambda^{i-1} u_{n-i}^2 + \lambda^m \sigma_{n-m}^2 \quad (8)$$

And by inspection equation (8) we can conclude that for large m the term $\lambda^m \sigma_{n-m}^2$ tends to zero and therefore equation (7) is the same as equation (6), where $\alpha_i = (1 - \lambda) \lambda^{i-1}$. In other words, weights decline exponentially at a rate λ as we move back through time.

Using a high λ , σ_n^2 responds slow to new daily returns reported by the S&P500 Index, and low λ responds fast. RiskMetrics database (JPMorgan 1994) used $\lambda = 0.94$ for updating daily volatility estimates. First we will use that number for λ to build a EWMA daily volatility model for the S&P500 Index, and then we try to choose the best λ which minimize the error between our EWMA model and the effective daily volatility observed in $t + 1$

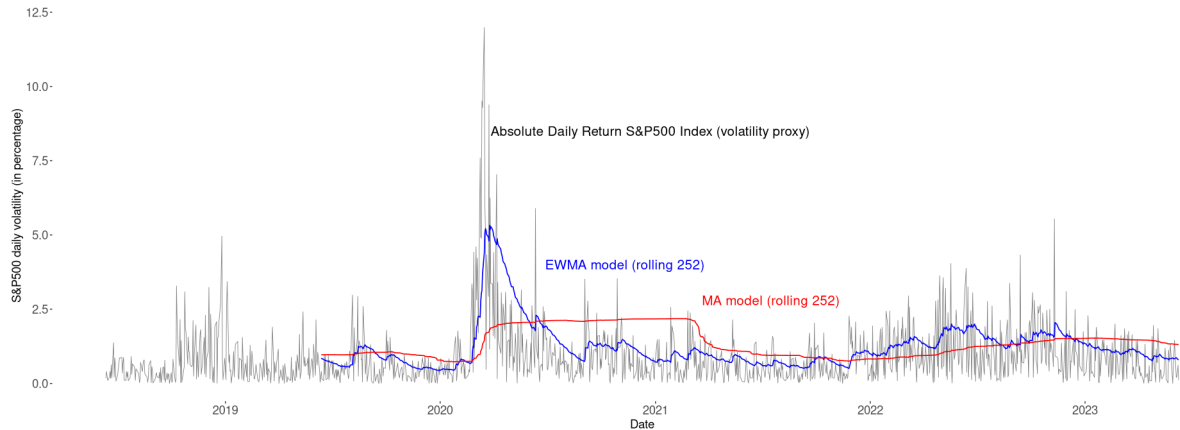


Figure 6: Basics weighting schemes volatility models: (1) EWMA model with a $\lambda = 0.94$ and (2) A simple MA model where all observations has the same weights. Both models use a window of 252 days.

As EWMA volatility model places more weight on recent data points ($\alpha_i = (1 - \lambda)\lambda^{i-1}$) versus a simple MA model which gives equal weight to all data points in the window ($\alpha_i = \frac{1}{m=252}$), the former one reacts more quickly to new information, and as showed Figure 6 has a better fit to forecast daily volatility than the simplest model (MA).

Exponentially weighted moving average models only depend on one parameter (λ). With the aim to calibrate this parameter, We will download a training set for the S&P500 Index form Yahoo Finance (11th June, 2013 to 11th June, 2018 - 5 years). Given the inherent unobservable nature of volatility, our proxy for the volatility will be the absolute daily return in percentage. By using this training set, we will test different values of λ and for each of those will measure how good the model fit the absolute return observed in $t + 1$ (proxy) using the Root Mean Square Error (RMSE) which is a measure that tells how far off our volatility predictions are from the absolute returns (proxy) on average.

$$\text{RMSE} = \sqrt{\sum_{i=1}^n \frac{(\hat{\sigma}_i - \sigma_i)^2}{n}} \quad (9)$$

where:

n is the number of observations

$\hat{\sigma}_i$ is the predicted daily volatility for day i

σ_i is the absolute return observed in day i (volatility proxy)

RMSE is a measure of the accuracy of a model in predicting quantitative data, in our case S&P500 daily volatility. Lower values indicate better fit, as there is less residual variation.

By using our training dataset, we calibrate the parameter λ which minimize the prediction error using RMSE and visualize if this optimum is significantly different than the lambda proposed by JPMorgan in 1994.

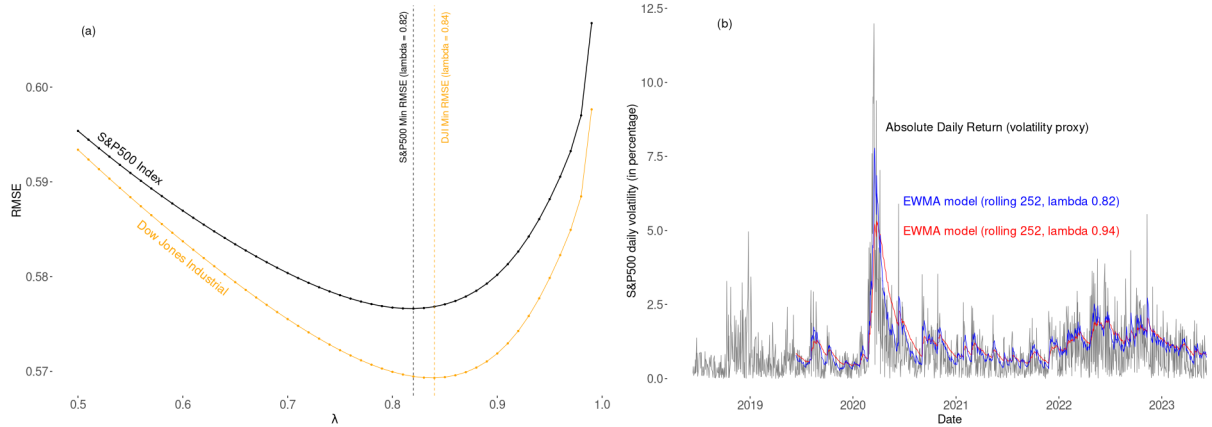


Figure 7: (a) RMSE for different EWMA volatility model as a function of λ using the training dataset and (b) Contrast two EWMA models: using a $\lambda = 0.94$ versus the optimum λ which minimize the RMSE ($\lambda = 0.82$)

Figure 7(a) reveals that the optimal value for lambda that minimizes fitting error using the training dataset (measured by the RMSE) is 0.82 for the S&P500 Index, which was in line with the optimum lambda for the Dow Jones Industrial (0.84). These optimum lambdas were lower than the lambda value of 0.94 proposed by JPMorgan in 1994. Lower lambda values enable the EWMA volatility model to respond more rapidly to new daily returns. This rapid response is crucial when forecasting over shorter horizons, like our focus on 1-day forecasts, where recent returns play a significant role.

Moreover, this characteristic of lower lambda values has shown to be particularly beneficial in the current market environment, which is experiencing high levels of volatility due to the impacts of COVID-19. By

reducing the lambda value, the model increases the weights placed on more recent observations, allowing it to capture unexpected shocks with greater accuracy.

Our training dataset suggests a quadratic relationship between the RMSE and the lambda parameter. The fitting error produced by $\lambda = 0.94$ is comparable to that produced by a $\lambda = 0.61$ equals to 0.586. However, these error levels are notably higher than the error level of 0.577 obtained using the optimal $\lambda = 0.82$.

Figure 7(b) further supports this conclusion by demonstrating how the optimal lambda improves the accuracy of the EWMA volatility model during periods of drastic returns volatility. A prime example of this occurred at the onset of the COVID-19 crisis in early 2020, where the chosen lower lambda value effectively captured the sharp spikes in volatility.

Now as we have our best EWMA volatility model (rolling 252 days, $\lambda = 0.82$), by using equation 5 and the prediction of volatility ($\hat{\sigma}_n$), we can estimate a Value-At-Risk for the S&P500 Index as shown Figure 8.

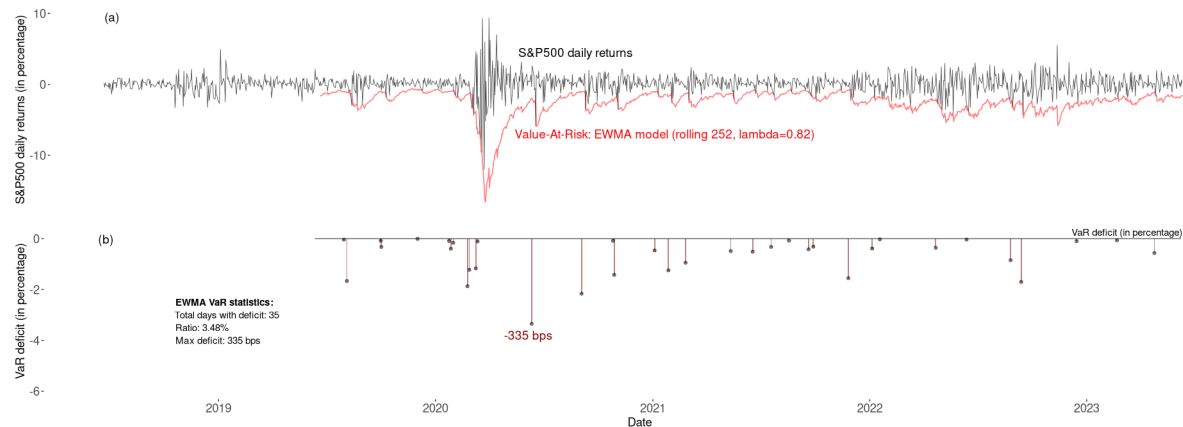


Figure 8: **(a)** Value-At-Risk using a EWMA volatility model vs realized daily returns and **(b)** VaR deficit: maximum loss estimated by the model (VaR) underestimate the negative daily return observed (difference in percentage)

In the analysis of our VaR using a EWMA model, it was found to underestimate the maximum loss, measured as negative daily return in percentage, on 35 out of the 1,005 days considered. This represents approximately 3.48% of the total days. The most substantial deficit occurred on June 11th, 2020, when the model predicted a maximum loss of 2.54%, while the observed negative return was 5.89%. This discrepancy, a shortfall of 335 basis points, is illustrated in Figure 8(b).

It is important to highlight that the deficit on June 11th, 2020 was influenced by significant macroeconomic events. The Federal Reserve had a meeting a day earlier, where they decided to keep interest rates near zero through 2022 due to the economic damage caused by the pandemic. The Federal Open Market Committee (FOMC) also predicted a 6.5% contraction in real GDP in 2020, with an unemployment rate at 9.3%. Additionally, the bearish remarks made by the Federal Reserve Chairman Jerome Powell, which presented a reality that the market did not want to hear, may have contributed to the negative market reaction.

Nevertheless, compared to a simple Moving Average (MA) model, our EWMA model demonstrates a substantial improvement in forecasting daily volatility. This improved performance can be attributed to the capability of the Exponentially Weighted Moving Average model to adapt more swiftly to recent market changes, a feature that proves especially advantageous in periods of sudden volatility spikes, such as those witnessed during the COVID-19 pandemic.

Yet, the EWMA model, while effective, is relatively simplistic in its weighting scheme, primarily distinguishing between more recent and older observations. In the following section, we will explore the application of the GARCH(1,1) model to forecast the S&P 500's daily volatility. It will be interesting to see how its performance compares with that of the EWMA model in the context of our data.

Model 2: Generalized Auto-regressive Conditional Heteroskedasticity (GARCH(1,1)) ⁴

An extension of equation 6 is to assume that exists a long-run average variance term (V_L) with a given weight, which leads to a GARCH(1,1) that takes the form:

$$\begin{aligned} \text{GARCH}(1,1) \Rightarrow \sigma_n^2 &= \gamma V_L + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 \\ \text{s.a: } \gamma + \alpha + \beta &= 1 \end{aligned} \quad (10)$$

where:

(1,1) means that σ_n^2 depends on most recent return observed (u_{n-1}) and most recent estimate of the variance rate (σ_{n-1}^2)

By inspecting equation 10 above, we can observe that the main distinction between our previous volatility model (EWMA) and a GARCH(1,1) model lies in the inclusion of a long-run average variance term with a specific weight in the formula (denoted as γV_L). Consequently, we can consider the volatility model developed in the previous section as a specific instance of a GARCH(1,1) model, where the weights used are: $\gamma = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$.

In the EWMA model, the long-run average variance term is not explicitly included, resulting in a simpler weighting scheme that primarily distinguishes between more recent and older observations. However, in the GARCH(1,1) model, the inclusion of the long-run average variance term allows for a more comprehensive representation of volatility dynamics.

The weights γ , α , and β in the GARCH(1,1) model control the influence of the long-run average variance, the most recent return observed, and the most recent estimate of the variance rate, respectively. By adjusting these weights, the model can capture both short-term volatility changes and the persistence of volatility over time.

In a GARCH(1,1) model used to predict daily volatility, the model's inherent stability is premised on a critical constraint: sum of the coefficients for the most recent return observed (α) and the most recent estimate of the variance rate (β) should not exceed 1 ($\alpha + \beta < 1$). A violation of this constraint would suggest that the weight assigned to the long-run average variance (γ) would be negative, which is not conceptually sensible or meaningful in the context of volatility modeling. This is because negative variance is not a valid concept, thus it underscores the importance of this constraint in ensuring the model's viability and interpretability in the context of financial volatility forecasting.

Setting $\omega = \gamma V_L$, equation 10 can also be written

$$\begin{aligned} \text{GARCH}(1,1) \Rightarrow \sigma_n^2 &= \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 \\ \text{s.a: } \alpha + \beta &< 1 \text{ for the process to be stationary.} \end{aligned} \quad (11)$$

where:

$$V_L = \frac{\omega}{\gamma} \Rightarrow V_L = \frac{\omega}{(1-\alpha-\beta)}$$

Our subsequent step in the analysis involves the application of the Generalized Auto-regressive Conditional Heteroskedasticity (GARCH) model, specifically GARCH(1,1), to forecast daily volatility. To do this, we need to estimate the model's parameters, denoted as ω , α , and β . These are estimated by using Maximum Likelihood Estimation (MLE) method.

The MLE operates based on the idea of maximizing a likelihood function. This function captures the probability of observing the given data under the chosen model parameters. For the GARCH model, the

⁴Bollerslev, T. (1986) 'Generalized autoregressive conditional heteroskedasticity', Journal of Econometrics, 31(3), pp. 307-327.

likelihood function assumes that the standardized residuals, which are the raw residuals divided by the estimated conditional standard deviation, follow a specific distribution.

At the outset of our analysis, we assume these standardized residuals follow a Normal distribution, which means that the probability distribution of daily returns (u_i) conditional on the variance is normal and therefore the likelihood function is:

$$\prod_{i=1}^m \left[\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-u_i^2}{2\sigma_i^2}\right) \right]$$

Taking logarithms, we get the log-likelihood function for a GARCH(1,1) model:

$$\sum_{i=1}^m \left[\ln(\sigma_i^2) - \left(\frac{-u_i^2}{2\sigma_i^2} \right) \right] \quad (12)$$

The weights are found by maximizing the log-likelihood function. This is typically done using numerical methods⁵ such as gradient descent or the Newton-Raphson method. The algorithm starts with initial estimates for the parameters, and then iteratively adjusts them to increase the log-likelihood until it reaches a global maximum.

In our work we will use a rolling window of 252 days, with the aim to dynamically update the model parameters over time. This involves re-estimating the model parameters every day, using the most recent 252 days of data. This approach assumes that the most recent data is the most relevant for forecasting future volatility.

In practice, this means that our m in the log-likelihood function defined in equation 12 will be 252, and we would take the preceding 252 days of returns, fit a GARCH(1,1) model to this data to estimate the parameters (ω , α , and β), then use these parameters to forecast the next day's volatility. Then move one day forward and repeat the process. This allows the model to adapt to changing market conditions, as we will see later by inspect the evolution of these optimum weights on time.

A GARCH(1,1) volatility model and the evolution of the weights can be interpreted in three phases by using Figure 9 as reference:

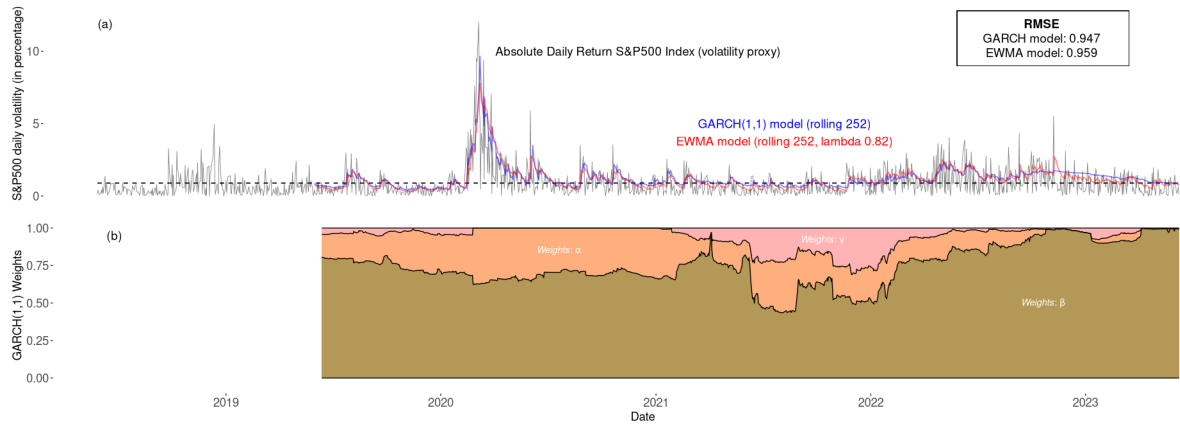


Figure 9: **(a)** Daily volatility of the S&P500 Index modeled by a GARCH(1,1) process and a EWMA (lambda 0.82). **(b)** Estimation of the GARCH(1,1) model's parameters (α , β , and γ), computed using Maximum Likelihood Estimation (MLE) - 252-day rolling window.

⁵Fiorentini, G., Calzolari, G. & Panattoni, L., 1996. Analytic Derivatives and the Computation of Garch Estimates. Journal of Applied Econometrics.

- (1) **Drastic Volatility Increase** (March 2020 - January 2021): During this period, characterized by a significant increase in observed volatility due to the COVID-19 crisis, the weights associated with the long-term variance (V_L) were estimated as zero. This indicates that the model converges to an EWMA model with a dynamically adjusted lambda. In this case, lambda is represented by the beta parameter, which fluctuated between 0.62 and 0.73 during this period. These values are lower than the optimum λ of 0.79 estimated in Model 1 (EWMA) and markedly lower than the lambda proposed by JPMorgan in 1994. This suggests that in periods of unexpectedly high volatility, these models tend to reduce λ values to assign more weight to recent observations, enabling them to capture short-term volatility spikes more accurately.
- (2) **Calm Market** (February 2021 - May 2022): In contrast, during periods of stability in the S&P500 market, when observed volatility aligns with the long-term average (as indicated by the red dashed line in Figure 9(b)), the GARCH(1,1) model significantly increases the weights assigned to the long-term variance (γ). This adjustment helps the predicted volatility converge more quickly towards the long-term average, thereby stabilizing the model.
- (3) **Slight Volatility Increase** (June 2022 - June 2023): In the final phase, where volatility shows a slight but persistent increase, the GARCH(1,1) model significantly reduces the weight given to the long-term variance ($\gamma \approx 0$) and considerably increases the beta values. This change is designed to put more emphasis on the most recent estimate of the variance rate (σ_{n-1}^2). During this period, the beta fluctuated between 1.0 and 0.83, with an average value of 0.93. This was greater than the beta values observed during the periods of extreme volatility and calm markets.

When evaluating the effectiveness of volatility prediction, the GARCH(1,1) model displayed superior performance compared to the previously model analyzed on this work (EWMA, lambda 0.82). This assessment was conducted using the Root Mean Square Error (RMSE), where the error using a GARCH(1,1) was 0.947 which was lower than the error reported using a EWMA model (0.959). Based on these findings, there is a strong suggestion that the GARCH(1,1) model could provide a more precise forecast of daily Value-At-Risk for the S&P500 Index.

By using Maximum Likelihood Estimation (MLE) to estimate the parameters α , β , and ω , it's evident that the GARCH(1,1) volatility model can effectively capture the dynamics of volatility clustering that are characteristic of various market conditions, such as turbulent, slightly volatile, or calm markets. However, a crucial question remains: how well does the model fit the underlying data? More specifically, does the assumption that the standardized residuals, which are integral to our chosen likelihood function, follow a normal distribution hold true in reality? This assumption is fundamental to the reliability of our GARCH model's volatility forecasts and must be thoroughly examined and validated.

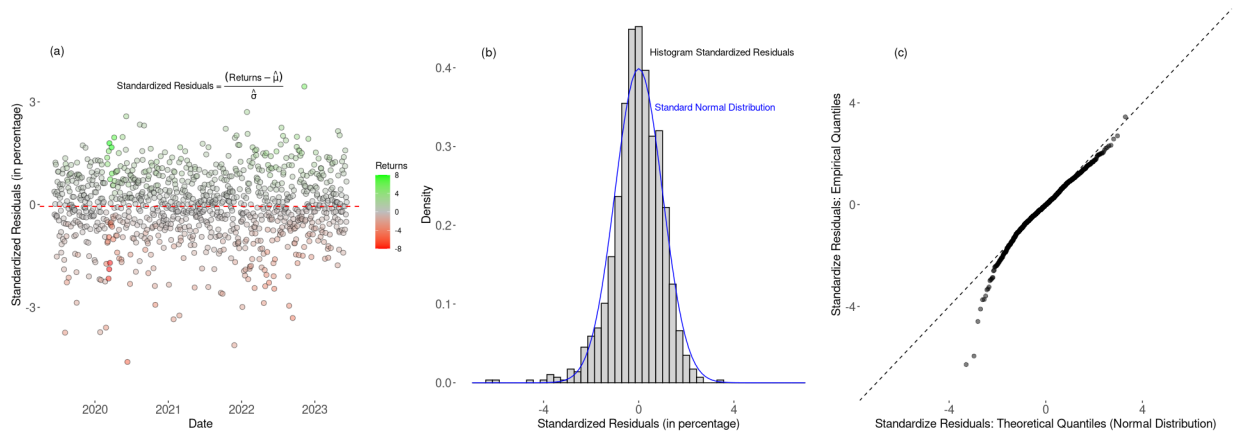


Figure 10: **(a)** Standardized residuals GARCH(1,1) volatility model, **(b)** Histogram of standardize residuals GARCH(1,1) model vs Standard Normal Distribution and **(c)** Theoretical quantiles for a Normal Distribution vs Empirical quantiles Standardize Residuals GARCH(1,1) model

In a GARCH(1,1) model, standardized residuals are defined as the difference between observed returns and

the model's estimated mean of returns, divided by the model's estimated volatility.

In an effort to elucidate the empirical standardized residuals of our volatility models, we adopt a strategy of selecting the last standardized residual from each fitted window. Given that our data set comprises 1,257 daily returns and we fit our models using a window size of 252 days, this method generates a series of 1,005 standardized residuals for our standardized residuals analysis. By employing this methodology, we ensure a comprehensive assessment of standardized residuals across the full spectrum of volatility scenarios, such as turbulent, normal and calm financial markets.

Fig. 10(a) presents standardized residuals for a GARCH(1,1) volatility model. If the model is accurately defined for a standard normal distribution, points on this plot should scatter randomly around a zero mean on the horizontal axis. At first glance, the red-dashed line, representing the mean, approximates zero (≈ -0.05), with points randomly fluctuating around this mean over time. Despite this, the presence of several outliers suggests the need for further analysis to verify the assumption of normality in the GARCH(1,1) model's standardized residuals.

Our empirical study indicates that returns for the S&P500 are not normally distributed - a key consideration in the analysis of standardized residuals. To provide further insight, we color each point in our standardized residuals according to the return. Fig. 10(a) highlights that the most significant outliers occur when observed daily returns on the S&P500 are notably negative. This pattern leads us to suspect that the distribution of standardized residuals is not symmetrical. This asymmetry becomes more evident in Fig. 10(b), which shows a longer left tail - representative of high negative returns - compared to the right tail.

The QQ-plot illustrated in Fig. 10(c) further illustrates this asymmetry, with the left tail of standardized residuals significantly misaligned with the theoretical standard normal tail. In conclusion, various statistical measures and tests indicate that standardized residuals are not normally distributed. For instance, the Shapiro test rejects the null hypothesis of normal distribution, yielding a test-statistic of 0.96 with a p-value of $3.6e-15$ (less than 5%). Further, a skewness of -0.85 indicates an asymmetric distribution with a longer tail on the left side, supporting our suspicion of more extreme values in the lower tail.

Considering all these issues on our model, now using equation 5 and the prediction of volatility ($\hat{\sigma}_n$) provided by the GARCH(1,1) model, we can estimate a Value-At-Risk for the S&P500 Index as shown Figure 11.

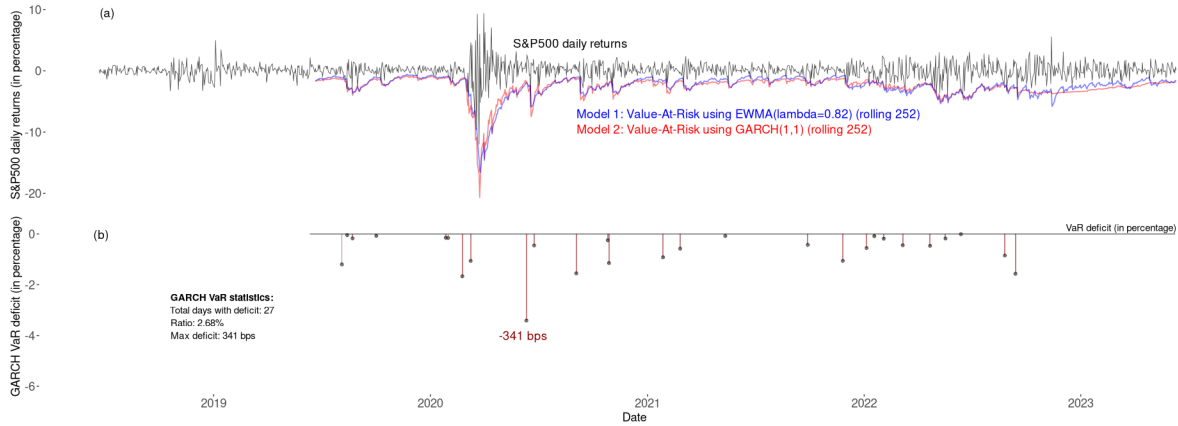


Figure 11: **(a)** Comparison between a Value-At-Risk using a GARCH(1,1) volatility model (red line) vs a EWMA(lambda=0.82) volatility model (blue line) and **(b)** Deficit between the VaR prediction using the GARCH(1,1) model and the observed daily return

As illustrated in Figure 11(a), the dynamic nature of the GARCH(1,1) model, which allows for fluctuating weights, enhances its ability to swiftly respond to transitions from stable to turbulent market conditions - a characteristic aptly demonstrated during the upheaval of the first half of 2020 induced by the pandemic crisis.

In scenarios of pronounced volatility spikes, GARCH(1,1) model strategically diminishes the weight it assigns to the long-run variance (γ), while simultaneously increasing the weight it places on the most recent observations. Conversely, during periods of relative market stability, the model readjusts to lend greater weight to the long-run variance, enabling a rapid convergence towards the long-term average volatility. This

degree of flexibility is a hallmark of the GARCH(1,1) model, enabling it to adapt quickly on different volatility market scenarios - a feat not mirrored by the EWMA model, which relies solely on a constant decay factor (λ) to encapsulate the dynamic of volatility.

When the market begins to oscillate at levels of volatility higher than the long-term average, as was observed for the S&P500 Index from June 2022 onwards, the GARCH models tend to provide more stable volatility predictions. This stability can be attributed to the model's strategic increase in the weights it assigns to the most recent estimate of the variance rate ($\hat{\sigma}_{n-1}$). In contrast, the EWMA models may display reduced accuracy, as their volatility forecasts in such market scenarios demonstrate higher variance than their GARCH counterparts.

In the context of forecasting a daily Value-At-Risk (VaR) using the GARCH(1,1) volatility model, it was found that the model underestimated the maximum loss (denoted as negative daily return in percentage) on 27 occasions out of the 1,006 predicted days. This represents an underestimation rate of approximately 2.7%. The most significant shortfall was noted on June 11th, 2020. On this day, the model predicted a maximum potential loss of 2.49%, while the actual observed negative return was 5.9%. This discrepancy, representing a shortfall of 341 basis points, is visually depicted in Figure 11(b).

While the statistical findings from the GARCH(1,1) model are not drastically different from those obtained using our initial EWMA volatility model (Model 1), there is a key distinction to be noted. The GARCH(1,1) model exhibits a higher degree of adaptability to sudden shifts in market conditions. It accomplishes this by dynamically adjusting the weights (γ , α , and β). This flexibility allows the GARCH(1,1) model to respond more promptly and accurately to drastic changes in the financial market environment, thereby potentially enhancing the reliability of its VaR predictions.

This updated volatility model has demonstrated several enhancements in comparison to our initial proposal, the Exponential Weighted Moving Average (EWMA). However, despite these advancements, the primary assumption of a GARCH(1,1) model - that standardized residuals should exhibit a standard normal distribution - was not fully met. Indeed, our findings illustrated a negatively skewed distribution in the standardized residuals, which deviates from the expected behavior.

Therefore, as we proceed with our analysis, we aim to develop a model capable of addressing these asymmetries in the empirical distribution of standardized residuals. Consequently, our focus will shift towards more advanced GARCH models, designed to handle asymmetric responses between returns and volatility (known as the leverage effect). The upcoming section will delve into the specifics of those robust models to forecast volatility, providing a comprehensive understanding of its utility in tackling the empirical distribution asymmetries observed on this section using a simple GARCH(1,1) volatility model.

5. Leverage Effect: Value-at-Risk by Addressing Asymmetric Distributions

While the last volatility model examined earlier in this work offers considerable flexibility in adapting to market changes through its dynamic weight adjustments, showed a substantial asymmetry in the distribution of standardized residuals. This discrepancy can be largely attributed to an inability to adequately account for the inverse relationship between asset returns and volatility, a phenomenon widely known as the *leverage effect*.

The main objective of this section is to delve deeper into the intricate dynamics of this leverage effect, to better understand the underlying causes that induce this phenomenon in financial asset prices. Drawing upon this understanding, we will endeavor to enhance the robustness of a simple GARCH(1,1) volatility model, as discussed in the preceding sections of this work, by incorporating the unique aspects of this inverse relationship.

At the end of this section, will propose an advanced GARCH(1,1) volatility model designed to handle the intricacies of the leverage effect. This model aspires to yield standardized residuals with a symmetric distribution, thereby improving the accuracy of volatility predictions. Through this enhancement, we hope to reconcile the theoretical assumptions of our model with the empirical realities of financial markets, ensuring that our volatility forecasting remains reliable and relevant under varying market conditions.

The Leverage Effect

In empirical terms, stock price volatility typically exhibits an inverse relationship with price. To illustrate, Fisher Black (1976)⁶ found that negative returns exert a greater influence on volatility than equivalent positive returns. This discovery of a leverage effect within the market component of volatility guides our subsequent analysis. Our intention is to ascertain the extent of this leverage effect within our dataset and evaluate the resulting asymmetric distribution between returns and volatility.

To conduct this analysis, first we need to segmented our dataset for the S&P500 by similar returns in absolute values with the aim to investigate that those returns were negatives consistently showed bigger future volatility on average than equivalent positive returns. We split our observations (daily returns) into four distinct categories, or ‘buckets’. Each bucket represents a range of daily absolute returns: the first spanning 0 to 1, the second 1 to 2, the third 2 to 3, and the final one covering 3 to 4. While our data does include absolute returns exceeding 4, these constitute less than 1.5% of our data, and we decided to focus the analysis across these four categories described before to avoid take any statistical inference by using extreme values of our data (outliers).

Having partitioned our dataset by bucket and further sorted each bucket by positive/negative returns, we aggregated these observations and computed the average volatility for the next day, and the average of the next five days’ volatility, which can provide us an opportunity to examine the persistence of the leverage effect throughout the week.

Our empirical investigation of the leverage effect present within the S&P500 Index is visually demonstrated in Figure 12. In these plots, the horizontal axis signifies the ‘buckets’ or groups into which we have categorized our data based on ranges of daily absolute returns. Conversely, the vertical axis represents the future daily absolute returns observable over subsequent days, serving as an approximate measure of forthcoming volatility.

Where these two charts diverge is in their different approaches used to measure the future volatility: Figure 12(a) presents 1-day forward volatility, while Figure 12(b) displays the average 5-day forward volatility. The latter is designed to explore whether the leverage effect is persistent over time.

To further decode the figure, each bucket in both plots features two distinct circles. The red circle represents the average future volatility observed for the negative returns within that particular bucket. In contrast, the green circle denotes the same measurement but for those instances where returns within the buckets were positive. This differentiation allows us to further examine and understand the leverage effect present within our S&P500 data set.

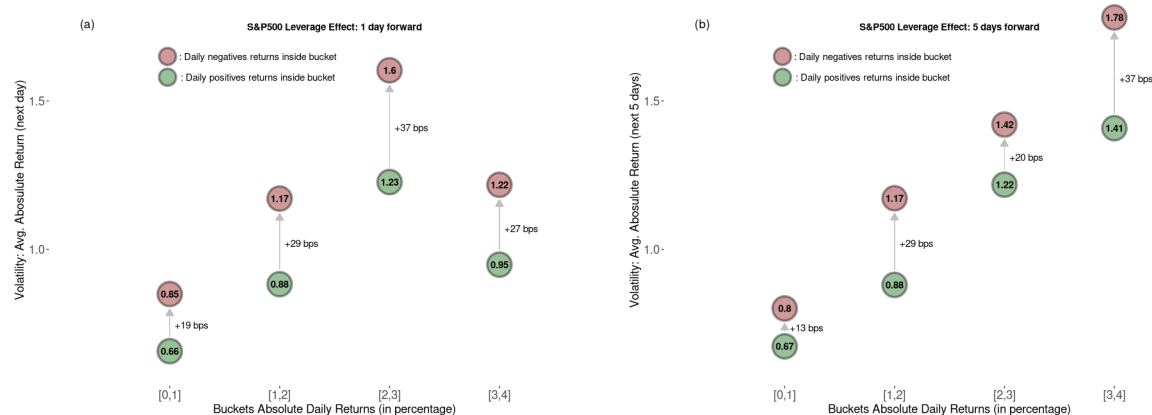


Figure 12: (a) Leverage Effect for the S&P500 Index using volatility as the next absolute daily return observed and (b) Leverage Effect for the S&P500 Index using volatility as the average of the next five absolute daily return observed

Consistent with Black’s findings, our results substantiate that negative returns exert a more substantial impact

⁶Black, F. (1976) ‘Studies of Stock Price Volatility Changes’, Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association, pp. 177-181.

on volatility than equivalent positive returns, thus highlighting an asymmetry in the relationship between returns and volatility. This asymmetric relationship contradicts our initial modelling using a GARCH(1,1), which presumed symmetric relationship. Figure 12(b) further underscores the temporal persistence of the leverage effect, showing that over the course of the subsequent five days, negative returns consistently exhibit higher volatility than their positive counterparts across all analyzed buckets.

The fundamental causes of the asymmetric relationship between returns and volatility, known as the leverage effect, can be traced back to the mindset of investors and their expectations. Every investor expects positive returns on their investments, while simultaneously acknowledging that the potential return is directly correlated with the associated risk. Consequently, understanding how investment risk is measured is crucial to identifying the potential causes of the leverage effect.

The value of a company's assets is a sum of its total debt and total equity⁷. This equation allows us to conceptualize a measure of risk that is directly related to the composition of these asset values.

$$\text{Assets}_{(t)} = \text{Total Debt}_{(t)} + \text{Total Equity}_{(t)}$$

Credit rating agencies meticulously analyze a particular ratio that gauges the proportion of the company's debt in relation to its own resources, or equity, known as the financial leverage ratio. An increase in this ratio raises the company's likelihood of bankruptcy, leading to a downgrade in credit rating by the agencies.

$$\text{Leverage Ratio}_{(t)} = \frac{\text{Total Debt}_{(t)}}{\text{Total Equity}_{(t)}}$$

For equity investors, the denominator of the leverage ratio is easily estimated because a proxy for equity value is the market capitalization, calculated as the product of the number of shares and the current market price.

$$\text{Total Equity}_{(t)} = \text{Number Shares}_{(t)} \cdot \text{Spot Price}_{(t)}$$

The concept of the leverage effect, as introduced by Fisher Black in 1976, suggests that a fall in the stock price directly translates into an increase in the leverage ratio. This rise in the ratio implies an increase in the risk associated with the investment, and consequently, the volatility. On the other hand, when the stock price rises, the debt-equity ratio declines, reducing the firm's financial risk and, consequently, the volatility of its equity. Thus, the key drivers of the leverage effect are the fluctuating market prices and their impact on the financial leverage ratio, influencing the perceived risk and resultant volatility.

Having delved into the intricacies of the leverage effect, we are now ready to progress to the final part of this section. This will encompass a comprehensive examination of volatility models within the GARCH family, specifically those that are designed to account for this significant phenomenon. Our investigation will include a comparative study, wherein these models will be evaluated in juxtaposition with the basic GARCH(1,1) volatility model previously formulated. The overarching aim is to critically analyze how incorporating the leverage effect can potentially enhance the predictive accuracy and robustness of these volatility models. Furthermore, we will examine the implications of these adjustments on VaR forecasting, thereby offering a more holistic understanding of the application of these models in financial risk management.

Model 3: GJR-GARCH(1,1) ⁸

This model was developed by Lawrence R. Glosten, Ravi Jagannathan, and David E. Runkle and improves the traditional GARCH(1,1) volatility model, by adding an additional term which becomes active when the lag return is negative ($u_{(t-1)} < 0$), enabling the model to account for the leverage effect.

⁷Lambert, R.A., 2012. Financial Literacy for Managers. University of Pennsylvania Press. (Chapter 'Your Company's Financial Health: What Financial Statements Can Tell You', p.33)

⁸Glosten, L.R., Jagannathan, R. and Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance, 48(5), pp.1779-1801.

An extension of equation 11 can be written as

$$\begin{aligned} \mathbf{GJR-GARCH(1,1)} \Rightarrow \sigma_n^2 &= \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 + \theta u_{n-1}^2 I_{(n-1)} \\ &= \mathbf{GARCH(1,1)} + \theta u_{n-1}^2 I_{(n-1)} \end{aligned} \quad (13)$$

where:

$$I_{(n-1)} = 1 \text{ when lag return was negative } (u_{n-1} < 0)$$

$$I_{(n-1)} = 0 \text{ when lag return was positive } (u_{n-1} \geq 0)$$

And the GJR-GARCH(1,1) model's parameters (ω , α , β and θ) are estimated by maximize the log-likelihood function described on equation 12

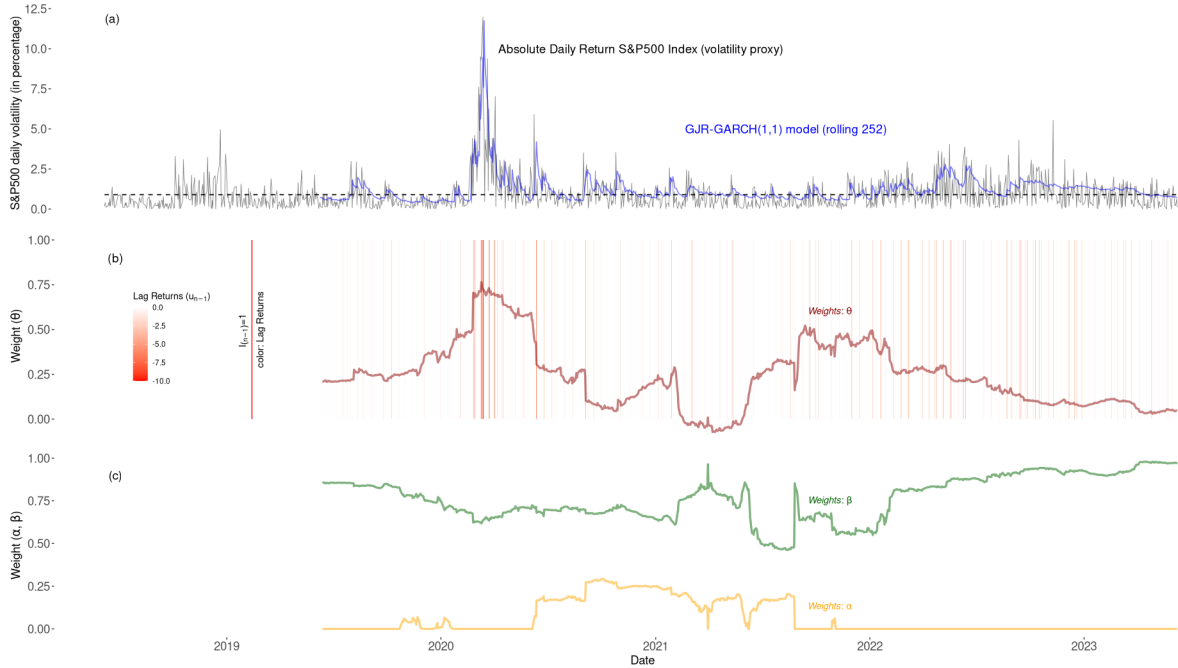


Figure 13: **(a)** Daily volatility of the S&P500 Index modeled by a GJR-GARCH(1,1) process. **(b)** Estimation of the GJR-GARCH(1,1) model parameter (θ), computed using Maximum Likelihood Estimation (MLE) - 252-day rolling window, and **(c)** Estimation of the GJR-GARCH(1,1) model's parameters (α and β), by the same methodology used to found θ (MLE)

The model's efficacy is highlighted during the market turbulence of the first quarter of 2020, triggered by the COVID-19 crisis. The GJR-GARCH(1,1) model effectively captures the sharp spike in absolute returns, as depicted in Figure 13(b). In this period, the weight for the leverage effect, theta, escalates to approximately 0.70. Concurrently, the activation of the dummy variable for negative lag returns ($I_{(n-1)} = 1$) amplifies the model's volatility, embodying the leverage effect extensively discussed earlier in this work.

Upon examining the other weights of the model (α and β) in Figure 13(c), and contrasting these with the parameters estimated using the classical GARCH(1,1) model as shown in Figure 9(b), no significant differences emerge. However, as we described earlier, the introduction of the new parameter (θ), could enhance the model's fitting.

Unfortunately as illustrates Figure 14, the lower tail of the standardize residuals using a GJR-GARCH(1,1) does not demonstrate an important improvement, and by looking those QQ-plots we cannot conclude which model adhere better the assumption of normality on the standardize residuals. The only conclusion we can make at this point is that the standardize residuals of both models are not normally distributed, and showed wider lower tails. To address this issue on the GJR-GARCH(1,1) model, we will employ two distinct

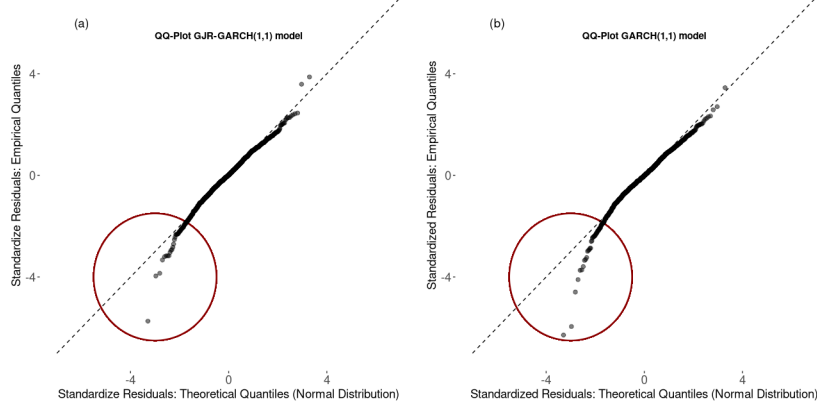


Figure 14: **(a)** Theoretical quantiles for a Normal Distribution vs Empirical quantiles Standardize Residuals GJR-GARCH(1,1) model, and **(b)** Theoretical quantiles for a Normal Distribution vs Empirical quantiles Standardize Residuals GARCH(1,1) model

methodologies; one assuming our current approach (normal distributed standardize residuals), while the other modifies the assumption to a Student's t-distribution with four degrees of freedom with the aim to model wider tails. Later we will select the best model by using a fitting performance criteria. The selected GJR-GARCH(1,1) model will then be utilized to forecast the next day's volatility, consequently estimating the VaR for the S&P500 Index.

How can measure numerically which model has the best fit based on our dataset? For this task we will use the Akaike Information Criterion (AIC), which is a measure used to compare the fit of different models to a given dataset. This statistic measurement helps in model selection by balance the complexity of the model, to avoid over-fitting, and its fits to the data.

$$\mathbf{AIC}_{(\text{python})} = 2k - 2\ln(\hat{L}) \quad (14)$$

$$\mathbf{AIC}_{(R)} = \frac{2k - 2\ln(\hat{L})}{N} \quad (15)$$

where:

k is the number of estimated parameters in the model (*complexity*)

\hat{L} is the maximum value of the log-likelihood function for the model (*by using MLE*)

N is the number of observations used to fit the model

The concept of utilizing the AIC as a fit performance metric to select between different volatility models will be central to our analysis. We will select the model that yields the lowest AIC value. Each time we fit a model with the intent of estimating optimal parameters through MLE, we will compute the AIC. Our models will be fit within a rolling window of 252 days, taking into account our entire dataset, which encompasses 1,258 daily returns (observations), this procedure will result in the calculation of 1,006 AIC values for each model.

By examining the series of AIC values, we will not only be able to determine which model exhibits the best fit but also track fluctuations in the AIC in relation to diverse volatility scenarios. This examination will enhance our comprehension of whether a particular model can more effectively capture specific market volatility patterns.

In practice, we apply three distinct volatility models: the conventional GARCH(1,1) model, and two GJR-GARCH(1,1) models. The former GJR-GARCH model operates under the assumption of normality for its standardized residuals, while the latter adopts a t-Student assumption with 4 degrees of freedom.

A comprehensive analysis of the fitting performance of these models is depicted in Figure 15. At the top, Figure 15(a), displays the daily volatility evolution, gauged via the absolute daily return of the S&P500 Index. To provide insight into average volatility within each 252-day window for model fitting, we include two additional lines on this plot: a red line representing the moving average of absolute daily returns observed over 252 days, and a dashed black line illustrating the average absolute daily return captured within the entire dataset. By juxtaposing these two lines, we can discern in which volatility scenarios the models demonstrated accuracy, and in which scenarios they underperformed.

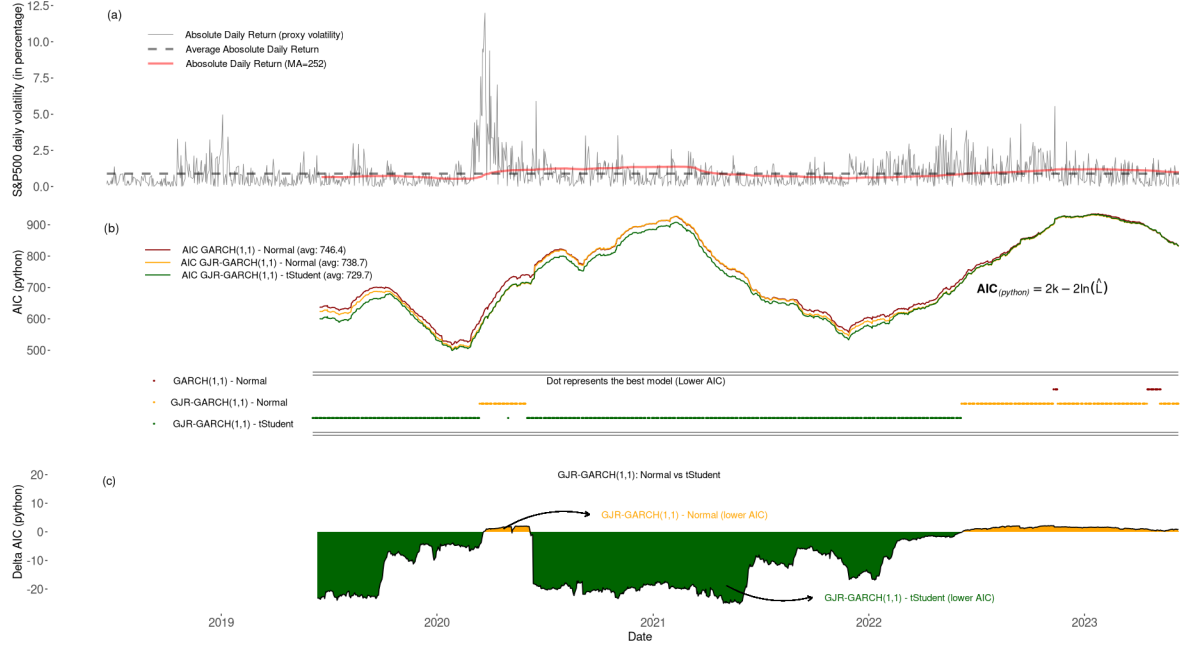


Figure 15: **(a)** Absolute daily return for the S&P500 Index (volatility proxy), **(b)** AIC scores for every 252 days window the models were fitting (lower scores, mean best fitting score), and **(c)** Difference on the AIC scores for both GJR-GARCH(1,1) models (Normal and t-Student) along time

As stated earlier, the AIC score is a robust statistical metric used for model selection. Figure 15(b) also indicates the frequency and circumstances in which each model presented the lowest AIC, implying the best fit for a specific fitting window. The GJR-GARCH(1,1) model consistently outperforms the traditional GARCH(1,1) model, which only demonstrated superior performance in 21 fitted windows (comprising 2.09% of the total). Further, the GJR-GARCH(1,1) model assuming a t-Student distribution for standardized residuals provided a superior fit in 701 windows (69.68% of the total), while the model assuming normality was preferred in the remaining 284 windows (28.23% of the total).

To elucidate the magnitude of differences between the two fitted GJR-GARCH(1,1) models, Figure 15(c) depicts a delta in the AIC scores. From this, it can be concluded that when the model assuming normality had the lowest AIC score (indicating the best fit), the differential in AIC scores remained minimal. However, as denoted by the green area on the chart, the GJR-GARCH(1,1) model with a t-Student assumption consistently exhibited a marked difference. By examining the collective average of the AIC scores (across 1.006 fitted windows), we note that the model accommodating wider tails (t-Student) delivered an average AIC of 729.7, surpassing the model presuming normality by a margin of 9 points (with an average AIC of 738.7) and outperforming the classical GARCH(1,1) model by a significant gap (16.7 points). This underlines the model's superior fit performance. Earlier, we spotlighted the persistent nature of the leverage effect over a subsequent five-day period. This finding is crucial as, notwithstanding the GJR-GARCH(1,1) presents inadequacy in accurately modelling the lower tail of the standardized residuals, it displayed a superior fit performance on average compared to a conventional GARCH(1,1) model. This attribute underscores the relevance of the GJR-GARCH(1,1) model, as evidenced by its lower average AIC score described before. The persistent influence of the leverage effect over a five-day period appears to be a key factor in this improved fit, thus favoring the GJR-GARCH(1,1) model over its conventional counterpart.

Within the GJR model family, the variant assuming a t-Student distribution for the standardized residuals yielded the most effective results. Based on this evaluation, the optimized GJR-GARCH(1,1) model, incorporating a t-Student assumption, will be employed to estimate future volatility. This model will serve as the foundation for our ensuing forecast of one-day ahead volatility, thereby facilitating the estimation of VaR for the S&P 500 Index, in accordance with equation 5.

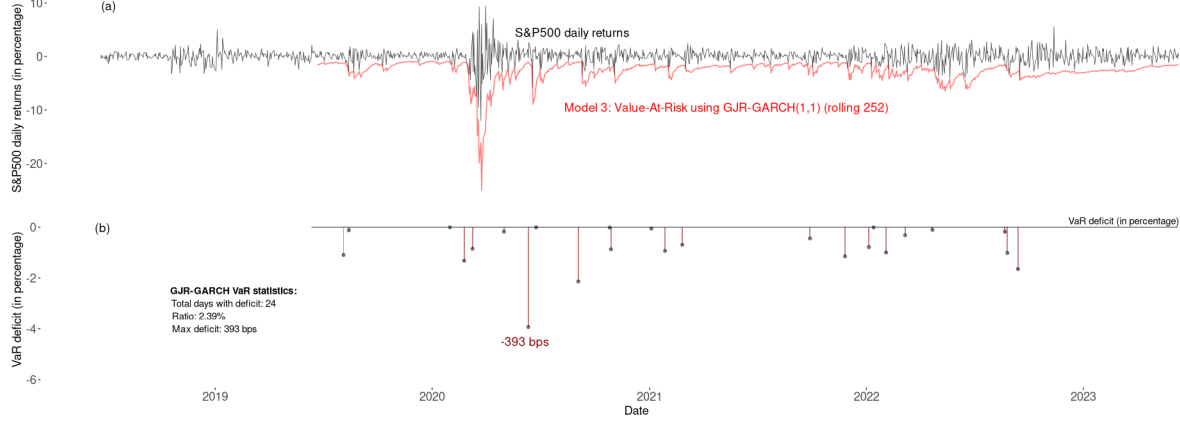


Figure 16: **(a)** Value-At-Risk using our best GJR-GARCH(1,1) volatility model (t-Student) and **(b)** Deficit between the VaR prediction and the observed daily return for the S&P500 Index

While the GJR-GARCH(1,1) volatility model showcased substantial improvements by consistently delivering lower AIC scores compared to the classical GARCH(1,1), its application to VaR modeling yielded conservative outcomes, as illustrated in Figures 16. The GJR-GARCH(1,1) underestimated the maximum loss (represented as negative daily returns in percentage) on 24 out of 1,006 predicted days. This represents an underestimation rate of approximately 2.4%. In an ‘ideal’ VaR model operating at a 95% confidence level, one would anticipate the maximum loss to exceed VaR estimates roughly 5% of the time. Despite our modeling efforts targeting the 5th percentile of the loss distribution, the observed maximum losses fell short of breaching this threshold.

Throughout the final months of 2022 and the initial semester of 2023, all models consistently exhibited comparable AIC scores. This suggests a level of parity in the model fit across the evaluated models. During this timeframe, none of the negative returns exceeded the VaR estimates derived from any of the models. The primary reason for this phenomenon could be attributed to the consistently minor absolute returns observed during this period, reflective of a calm market scenario. The coefficients of the models were fitted using a 252 days window of high volatility on average and this fit environment did not align with the prevailing market conditions as shown in Figure 15(a) where the moving average of absolute returns (252 days) was higher than the daily observed absolute return.

Moving forward, our research will explore another volatility model, the EGARCH(1,1), known for accommodating the leverage effect through an approach differing from our previously utilized model.

Model 4: Exponential GARCH Model, EGARCH(1,1) ⁹

In order to address the asymmetric relationship between returns and volatility, a new volatility model, known as the Exponential GARCH model (EGARCH), was introduced. This model, developed by Daniel Nelson¹⁰ in 1991, aimed to enhance the conventional GARCH models by effectively accounting for leverage effects. By modeling the logarithm of conditional variance as a function of past values and shocks, the EGARCH(1,1) model employs a distinctive approach to capture the asymmetric impact of positive and negative shocks on volatility, setting it apart from the GJR-GARCH model.

An extension of equation 11 can be written as

⁹Tsay, R.S. (2010) ‘3.8 The Exponential GARCH Model’, in Analysis of Financial Time Series. 3rd edn. Hoboken, NJ: John Wiley & Sons, Inc., pp. 124-130.

¹⁰Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica, 59(2), 347-370.

$$\text{EGARCH}(1,1) \Rightarrow \ln(\sigma_n^2) = \omega + \alpha \left(\frac{|u_{n-1}|}{\sqrt{\sigma_{n-1}^2}} \right) + \beta \ln(\sigma_{n-1}^2) + \theta \left(\frac{u_{n-1}}{\sqrt{\sigma_{n-1}^2}} \right) \quad (16)$$

where:

ω is a constant term

α captures the size of the shocks

β captures the persistence of volatility

θ captures the asymmetry or leverage effect

In our previous analysis employing the GJR-GARCH model, the mechanism to model the leverage effect was by used a dummy variable, denoted by $I_{(n-1)}$ in equation 13, which was triggered when the lag return was negative ($u_{n-1} < 0$). In contrast, the EGARCH model do not require a dummy variable to model the leverage effect and as depicted in equation 16, both positive and negative lag returns, exert an influence on volatility. To illustrate, if the parameter θ is negative, a negative lag return tends to inflate the volatility more than a positive. Furthermore, the EGARCH model does not require non-negativity constraints on its parameters like the GARCH model does. This is because on this particular model the variance, being the exponent of the model's terms, by definition is always positive.

By inspects equation 16, one discerns that the response of this new model is on a logarithm scale. This log-linear relationship between our response and explanatory variables injects an additional layer of complexity into the MLE optimization procedure when it attempts to locate a global maximum. To ensure convergence across all fitted windows, we adopted a grid-search approach, which facilitated the testing of various distributions for the standardized residuals to determine the one that best accommodates our data.

Employing a t-Student distribution did not yield satisfactory convergence, and the grid-search recommended increase the window size to over 424 days (data used to fit the model) or simplifying the assumption to a standard normal distribution with the aim to ensure convergence on the optimization procedure. We opted to fit the EGARCH(1,1) model using a standard normal distribution and not change our approach of using a 252-day window size.

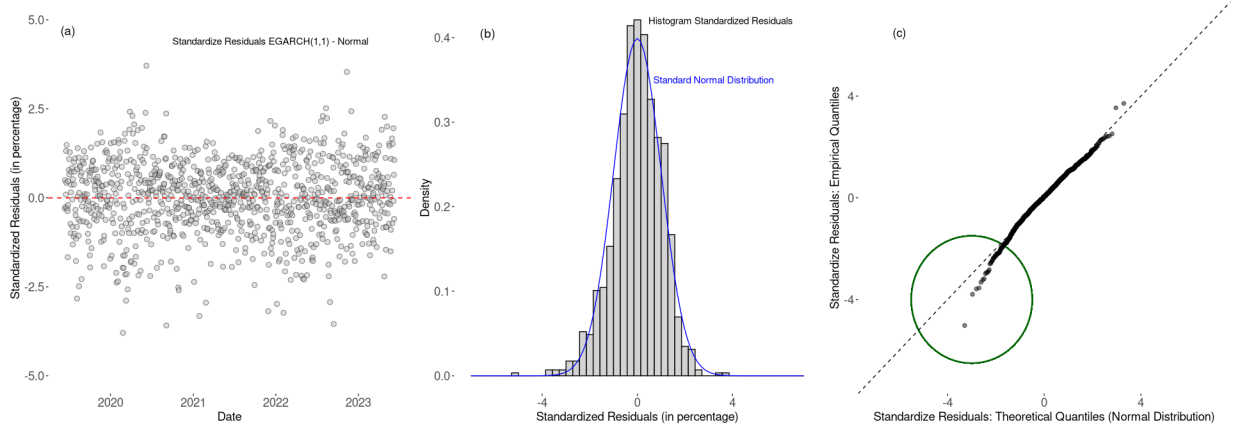


Figure 17: **(a)** Standardized residuals EGARCH(1,1) volatility model, **(b)** Histogram of standardize residuals EGARCH(1,1) model vs Standard Normal Distribution and **(c)** Theoretical quantiles for a Normal Distribution vs Empirical quantiles Standardize Residuals EGARCH(1,1) model.

The decision to assume a standard normal distribution appears justified when examining the behavior of the standardized residuals collected by using the methodology explained before (1,005 standardized residuals). Firstly, Figure 17(a) demonstrates that the standardized residuals exhibit a mean of zero (red dashed line), and

fluctuations around this mean appear to be random over time, devoid of any discernible patterns. Secondly, the contrast between a histogram of the empirical standardized residuals and the theoretical standard normal distribution (Figure 17(b)) reveals a good fit. Lastly, Figure 17(c) disclosed a minor deviation in the lower tail of the empirical distribution from the theoretical standard normal. Nevertheless, this disparity is considerably smaller than the one observed for the GJR-GARCH and conventional GARCH models. Consequently, the assumption of normality for the standardized residuals appears to be a reasonable one for the EGARCH volatility model and we hold this assumption to estimate our VaR for the S&P500 Index.

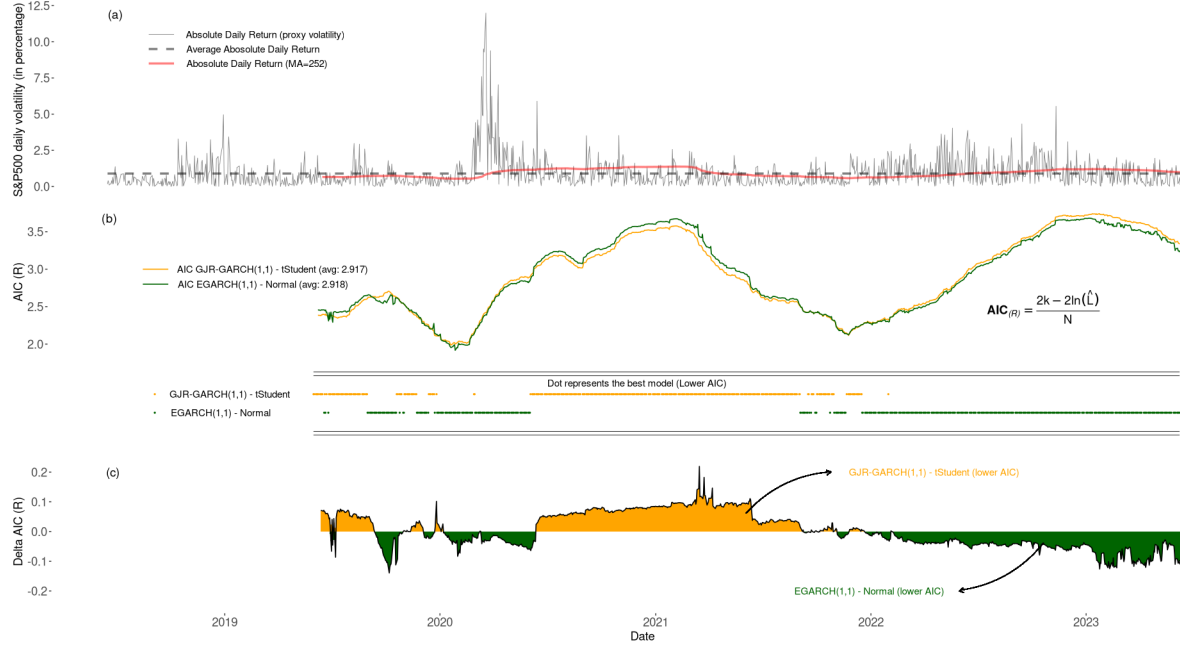


Figure 18: (a) Absolute daily return for the S&P500 Index (volatility proxy), (b) AIC scores for every 252 days window the models were fitting (lower scores, mean best fitting score), and (c) Difference on the AIC scores for a EGARCH(1,1) model vs a GJR-GARCH(1,1) model along time

The fitting performance of the EGARCH(1,1) volatility model is depicted in Figure 18. When we compare the two volatility models that account for the leverage effect, they yield similar AIC scores on average. However, a detailed examination of Figure 18(b and c) reveals notable variations in their performance across different time periods. For example, the EGARCH(1,1) model exhibited superior performance towards the end of our data series (green area). This is particularly interesting since, as previously discussed, the VaR estimates using the GJR-GARCH(1,1) model during this period were excessively conservative. In contrast, the GJR-GARCH(1,1) model demonstrated superior fitting performance at the end of 2020 and for most of 2021. As anticipated, both models, given their shared foundational premise of the leverage effect, exhibited a consistently superior fitting performance compared to the classic GARCH(1,1) model. However, the discrepancies in their approach to modeling the leverage effect led to significant variations in fitting performance during specific time periods.

Now, by using the optimal coefficients for the EGARCH(1,1) model obtained using a MLE approach, we can predict the next day's volatility ($t + 1$) and therefore estimate our VaR for the S&P500 Index using equation 3. These results are presented in Figure 19.

When the EGARCH(1,1) volatility model was used to estimate the 5th percentile of the daily loss distribution for the S&P500 Index (VaR), it underestimated the negative daily return 68 times, representing a rate of 6.76% (out of a total of 1,006 estimations). This rate displays a significant increase, specifically by 437 basis points, when compared to the GJR-GARCH(1,1) model. This comparison highlights that the EGARCH approach is less conservative, exceeding the desired 5% ratio consistent with an ideal VaR model at a 95% confidence level.

Contrasting the EGARCH(1,1) and GJR-GARCH(1,1) volatility models, a distinct difference in the VaR estimates becomes apparent. The EGARCH(1,1) model consistently displays more variability in its VaR

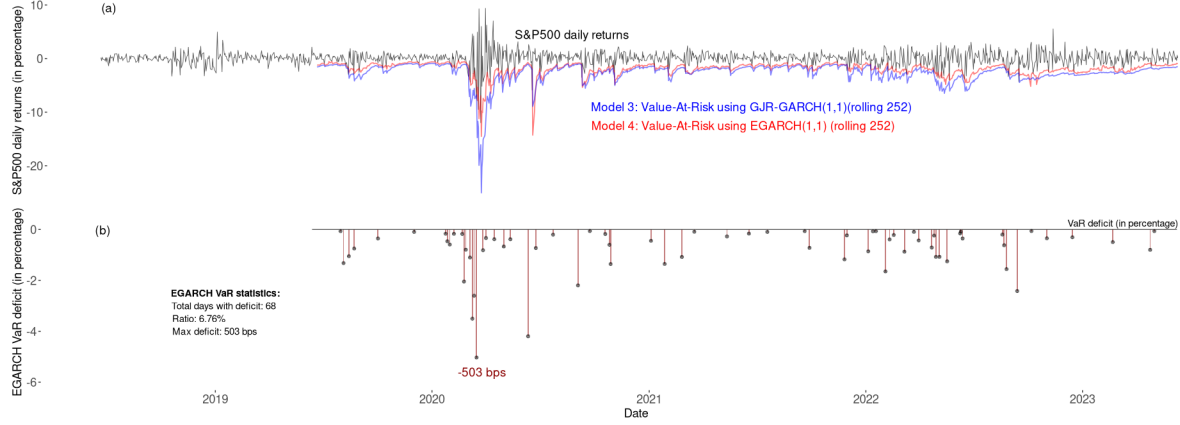


Figure 19: **(a)** Comparison between a Value-At-Risk using a EGARCH(1,1) volatility model (red line) vs a GJR-GARCH(1,1) volatility model (blue line) - both models accounts the leverage effect - and **(b)** Deficit between the VaR prediction using the EGARCH(1,1) model and the observed daily return

estimates compared to the GJR-GARCH(1,1) model. This higher fluctuations could provide an explanation as to why the EGARCH(1,1) model's estimations typically lean towards being less conservative than those of the GJR-GARCH(1,1) model.

An analysis of the VaR estimates from the final 12 months of our data set (June 2022 - June 2023) effectively demonstrates these differences. The GJR-GARCH model, represented by the blue line in Figure 11(a), reveals a relative stability in its VaR estimates throughout this period. Conversely, the EGARCH model, denoted by the red line, showed a similar trend but is characterized by more pronounced fluctuations. Furthermore, in this time frame, the GJR-GARCH(1,1) model exhibits high conservative VaR estimates, as evidenced by the absence of deficits. This stands in contrast to the EGARCH(1,1) volatility model, which underestimated the negative daily return on 13 occasions. These underestimations constitute a rate of 5.2%, calculated based on a total of 253 estimations, further reinforcing the less conservative nature and higher variability of the EGARCH(1,1) model's VaR estimates.

In the next section of this work, we will attempt to boost the volatility models analyzed by combining them, towards improving the VaR forecast for the S&P500 Index.

6. Boosting leverage effect volatility models

In the preceding sections, we conducted an in-depth analysis of four distinct non-constant volatility models with the primary objective of forecasting the fifth quantile of the loss distribution for the S&P500 Index, commonly referred to as the Value-At-Risk. Our initial investigations revolved around the empirical daily returns distribution, paving the way for the construction of our inaugural volatility models utilizing basics weighting schemes such as EWMA and GARCH processes.

However, it quickly became evident that these models fall short in encapsulating the asymmetric relationship between returns and volatility, a phenomenon often referred on the financial industry as the leverage effect. To address this problem, we introduced two advanced volatility models which, while being members of the GARCH family, also account for the leverage effect on their estimations. These models recognize the leverage effect differently, leading to substantial disparities in their fitting performance as measured by the AIC during certain periods.

Given these insights, the primary objective of this section is to merge these two models to elevate their potential and enhance our Value-at-Risk forecasting by harnessing the best aspects of both models.

Final adjustments GJR-GARCH(1,1): Optimum t-Student distribution (grid-search approach)

In the third volatility model analyzed, the GJR-GARCH(1,1), we assumed a t-Student distribution with 4 degrees of freedom for the standardized residuals and subsequently to estimate the 5th percentile of the

loss distribution (VaR). An important consideration we delve into is the appropriateness of choosing a fixed degree of freedom for every 252 days window where the model was fitted.

The t-Student distribution is characterized by a single parameter, specifically the number of degrees of freedom. It is noteworthy that the distribution's shape varies in accordance with this parameter. In instances where a small number of degrees of freedom is selected to fit the GJR-GARCH(1,1) volatility model, the resultant distribution will exhibit sharper peaks and heavier tails compared to a standard normal distribution. This implies an elevated likelihood of extreme values in a t-Student distribution, which in turn translates to a larger distance between the mean (μ) and the 5th percentile (lower tail).

Conversely, an increase in the number of degrees of freedom results in a decrease in the excess kurtosis of the t-Student distribution. For example, once the degrees of freedom surpasses 30, the shape of the t-Student distribution is virtually indistinguishable from that of a standard normal distribution. This transformation is visually depicted in Figure 20.

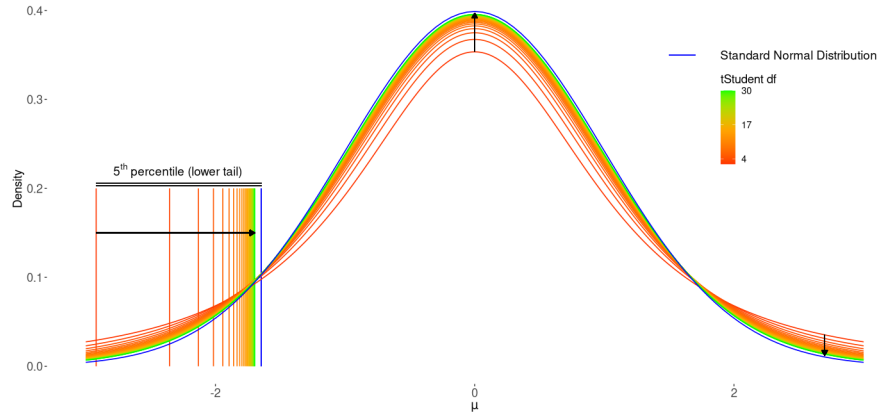


Figure 20: As degrees of freedom of a tStudent distribution increases, the shape of the distribution tends to a standard normal distribution

In the subsequent analysis, our attention pivots towards estimating the optimum degrees of freedom within each window where the model will be fitted. The prior section assumed a singular shape for the standardized residuals ($df=4$), which has an undefined kurtosis ($6/(df - 4) = 6/0$).

Given that each window's data varies, optimizing the log-likelihood through the estimation of a fifth parameter - the number of degrees of freedom - could be an effective strategy. This will enable us to choose the most suitable t-Student distribution shape, tailored to the specific data contained within each window.

To estimate the optimal number of degrees of freedom which maximize the log-likelihood function, we employed a grid-search methodology across the 1,006 windows where the GJR-GARCH(1,1) model was fitted. Moreover, to decrease the time consuming during the grid-search optimization, we restricted the range of the degrees of freedom to be between 3 and 50.

Results are shown in Figure 21. Interestingly, only 4.5% of the 1,006 windows where we applied the model resulted in an optimum degrees of freedom equal to 4 through the optimization procedure aimed at maximizing the log-likelihood function. The average optimum degrees of freedom across all windows where the GJR-GARCH(1,1) model was fitted was 12.52, notably higher than 4. Furthermore, in over 10% of the total windows fitted, the grid-search methodology suggested an optimal degrees of freedom for the t-Student distribution exceeding 30, thereby implying a distribution closely resembling a standard normal distribution.

From these findings, it can be inferred that the conservative VaR estimates (GJR-GARCH) could potentially be attributed to an overestimation of tail events. This is likely due to using a smaller degrees of freedom than the optimum number, leading to a wider tail distribution, and therefore a higher VaR predictions.

Now as we have the optimum degrees of freedom for the t-Student distribution, we can rewrite equation 5 as:

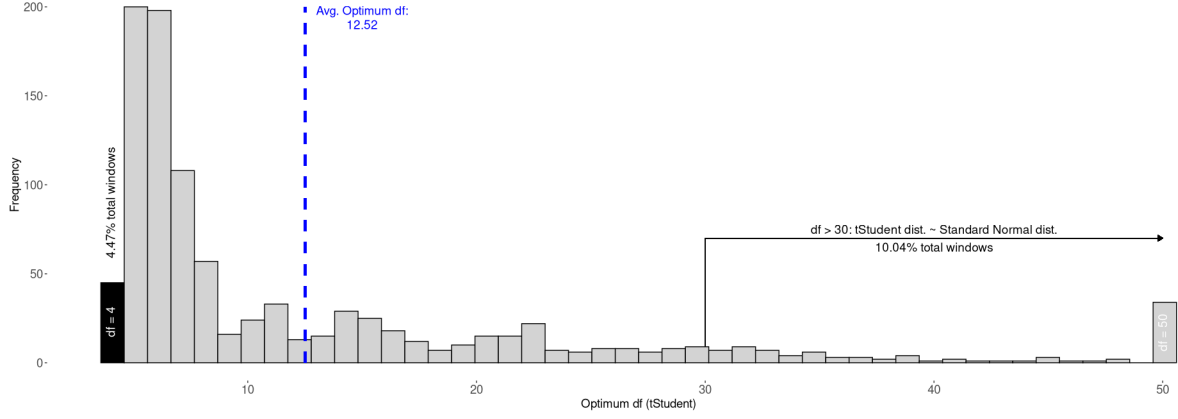


Figure 21: Histogram for the results of the grid-search approach to find the optimum degrees of freedom which maximize the log-likelihood function on each window where the GJR-GARCH model was fitted (1,006 total windows)

$$\mathbf{VaR}_{(0.95, 1_{day})} = \hat{\sigma}_{(1_{day})} \cdot t_{(0.95, df=df^*)} \quad (17)$$

where:

df^* represents the optimum degrees of freedom.

By applying the daily volatility forecast generated by our enhanced GJR-GARCH(1,1) model to equation 17, we are able to estimate the Value-At-Risk. This improved volatility model is referred to as *Model 3 v2* in acknowledgment of its evolution from the earlier developed Model 3. The resultant VaR estimates from this enhanced model is visually represented in Figure 22.

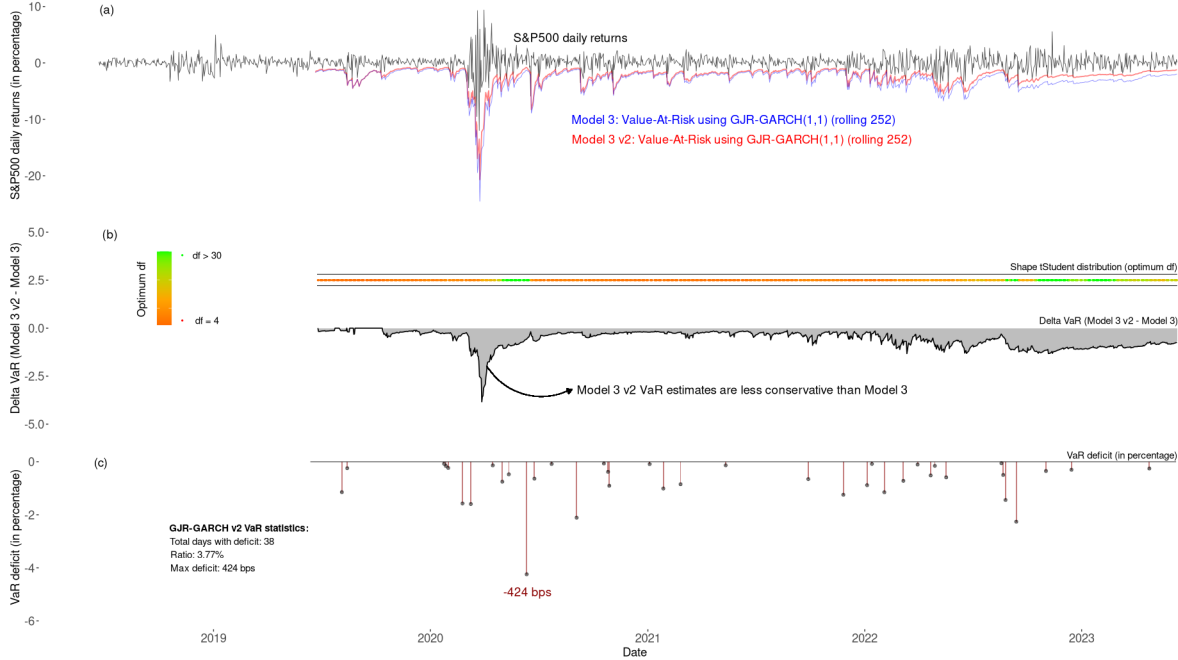


Figure 22: (a) Comparison between a Value-At-Risk of a GJR-GARCH(1,1) volatility model using optimum tStudent distribution (Model 3 v2) vs a GJR-GARCH(1,1) volatility model using 4 degrees of freedom on all windows the model was fitted (Model 3) - both models accounts the leverage effect - (b) Delta Value-At-risk estimates between Model 3 v2 and Model 3, and (c) Deficit between the VaR prediction using Model 3 v2 and the observed daily return

It is evident that when the optimum degrees of freedom for the t-Student distribution, in a specific 252-day window, was estimated as a small number (ranging between 4 and 10), the gap between the maximum loss predicted by Model 3 and the ‘improved’ model (Model 3 v2) tends to be minimal as shown Figure 22(b). However, when the optimum degrees of freedom for the t-Student distribution surpasses 30 (depicted as green dots), the discrepancy between the VaR estimates of the two models substantially enlarges. This indicates that under these conditions, Model 3 v2 - the advanced version - presents a less conservative maximum loss estimation for the S&P500 daily returns.

Assessing instances where the daily maximum loss predicted using Model 3 v2 underestimated the negative daily return of the S&P500 Index reveals important information. Out of 1,006 daily VaR forecasts, Model 3 v2 under-projected the actual loss 38 times. This led to a deficit ratio of 3.77%, a figure notably higher than that of the original GJR-GARCH(1,1) model, developed earlier in this study (+138 bps). Despite Model 3 v2 being less conservative in its estimations, the deficit ratio still falls short by 137 bps of the ‘ideal model’ to estimate the 5th percentile of the loss distribution.

Finally, as we now have two distinct well develop volatility models - both accounting for the leverage effect using different methodologies - our next step will be to construct a final VaR model. This model will dynamically select the superior estimate from both models within every window, with the objective of enhancing the precision of our VaR estimates.

Model 5: The best of both worlds (GJR-GARCH and EGARCH)

The fifth and culminating volatility model in this work is designed to merge the aspects of the two preceding models (Model 3 v2 and Model 4) that accounts for the leverage effect. The merger of these models will be conducted over a total of 1,006 fitted windows (252 days), during each of which we will utilize the Akaike Information Criterion (AIC) as the fitting performance metric to determine the superior model and estimate the VaR for the S&P500 Index as follows:

$$\begin{aligned} \Delta AIC_{(t)} &= AIC_{\text{Model 4 (t)}} - AIC_{\text{Model 3 v2 (t)}} \\ \text{VaR}_{(t+1)} &\begin{cases} \text{Model 3 v2: GJR-GARCH} & \frac{\Delta AIC_{(t)}}{AIC_{\text{Model 3 v2 (t)}}} \geq \text{threshold}, \\ \text{Model 4: EGARCH} & \frac{-\Delta AIC_{(t)}}{AIC_{\text{Model 4 (t)}}} \geq \text{threshold}, \\ \text{Average Models}^{11} & \text{Otherwise} \end{cases} \end{aligned}$$

Given that both volatility models embody the leverage effect through distinct methodologies, a threshold of 3% will be established for estimating the Value at Risk (VaR). In situations where the AIC differential between both volatility models falls below 3% of the lowest AIC, the models are deemed similarly effective. In such cases, rather than selecting one model over the other, we will employ the average of both models for the VaR estimations.

The thresholds and Delta AIC for each window wherein the models were fitted are illustrated in Figure 23. Intriguingly, in 92% of these instances, the $\Delta AIC_{(t)}$ did not surpass the established threshold. This finding translates to the majority of VaR estimations being derived from an averaged of both models, as the fitting performance differences between the two were not substantial enough to warrant a single model selection.

In the pursuit of forecasting the VaR for the S&P500 Index, the final model - designated as Model 5 - predominantly averages the outputs of both volatility models. Consequently, the VaR estimates produced by this model exhibit a level of conservatism that is intermediate between the two prior models used in this study - the EGARCH(1,1) and the improved GJR-GARCH(1,1). In other words, Model 5’s estimates are less

¹¹means averaging the VaR estimated from the two models (Model 3 v2 and Model 4)

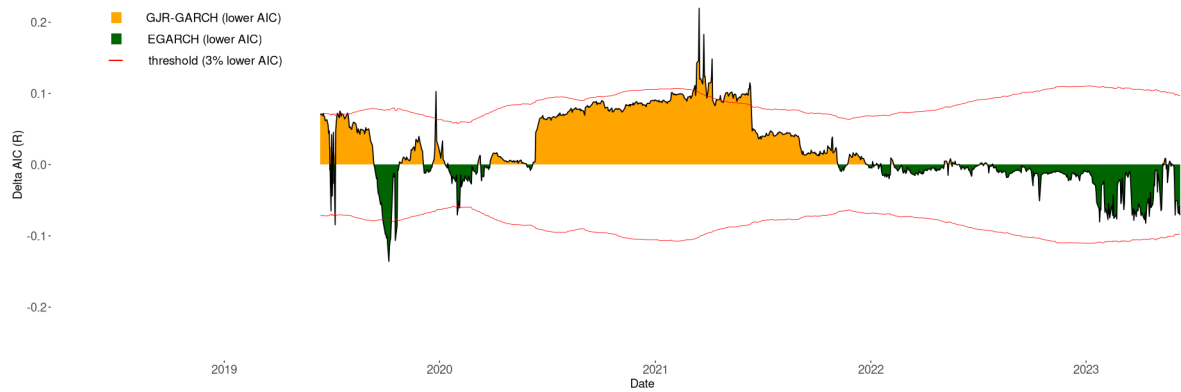


Figure 23: Difference on the AIC scores for GJR-GARCH(1,1) improved model vs EGARCH(1,1) model model along time. Red dashed line is the 3% threshold of the lowest AIC between both models

conservative than those derived from the GJR-GARCH(1,1) model, but conversely more conservative than the EGARCH(1,1) model. Illustrations of the VaR estimations, particularly at the 5th percentile of the loss distribution for the S&P500 Index, are presented in Figure 24. This distinctive approach of Model 5, which opts for a balance rather than favoring a specific volatility model, is attributed to the model selection criteria during the fitting process.

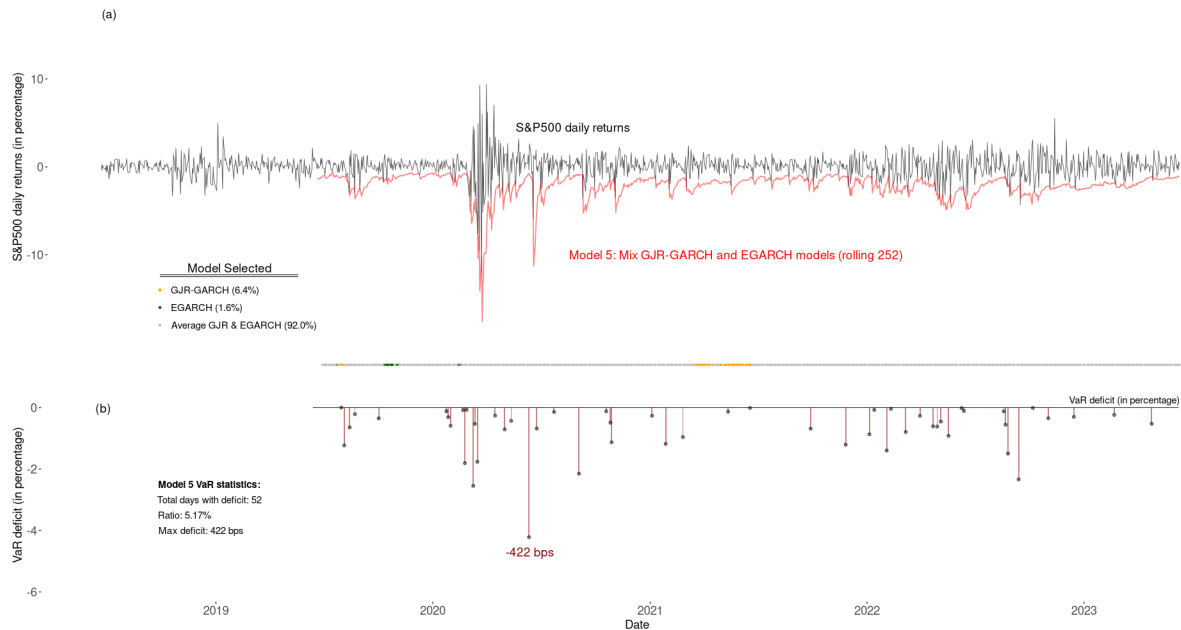


Figure 24: (a) Value-At-Risk estimated by using a mix volatility model (GJR-GARCH and EGARCH) and (b) Deficit between the VaR prediction and the observed daily return for the S&P500 Index

Evaluating instances where Model 5's prediction of daily maximum loss fell short of the actual negative daily return of the S&P500 Index, we find a deficit ratio of 5.17% (52 out of 1,006 total). This ratio closely aligns with the 'ideal' model prediction for the 5th percentile. Interestingly, the maximum deficit occurred on June 11, 2020 (422 bps), a result heavily influenced by the Federal Reserve's unexpected decision to maintain near-zero interest rates through 2022 due to pandemic-related economic damage. This unforeseen move triggered a negative market reaction on that specific day. As stock prices follow a stochastic process, significant movements driven by such macroeconomic events can lead to an underestimation of maximum loss by the VaR models used. This level of deficit becomes particularly concerning when estimating the Expected

Shortfall.

VaR estimates, while providing a maximum loss threshold with a given level of confidence (e.g., 95% of the time), fail to provide insights into the potential magnitude of loss once that VaR limit is exceeded. Expected Shortfall helps address this, offering an estimate of the expected loss when a loss surpasses the VaR threshold. It essentially represents the average of all the losses that exceed the VaR, thereby providing more information about the potential ‘tail risk’ or the risk of extreme negative events. However, the primary focus of this study remains on the VaR estimates derived from five different volatility models developed.

In the subsequent section of this report, we will back-test the five VaR estimates, a crucial process to effective financial risk management. The main objective of back-testing is to evaluate the accuracy and reliability of the VaR model, comparing its forecasted loss distributions against the actual losses occurring over a historical period.

————— From here Valentin —————

7. Back-Testing: Assess models performance

The goal of this section is to rigorously evaluate the precision of the five volatility models developed in this work by counting the number of instances where the day’s negative return exceeds the one-day 95% Value-At-Risk as predicted by the respective model. Such instances, where the actual loss surpasses the VaR estimated, are identified as *exceptions*.

In the preceding sections of this work, we quantified these exceptions as VaR deficits. We tallied the number of exceptions and calculated a deficit ratio, defined as the number of exceptions divided by the total number of predictions. This ratio was subsequently compared to that of an ‘ideal’ VaR Model at a 95% confidence level, which should theoretically yield a deficit ratio of five percent. If a specific VaR model displayed a deficit ratio below 5%, we inferred the model to be more conservative since it underestimated the ‘ideal’ model. Conversely, if the deficit ratio exceeded 5%, we concluded the model to be less conservative because it overestimated the ‘ideal’ model.

In this section, we will employ a formal statistical test method to assess the performance of the model to estimate the 5th percentile of the loss distribution for the S&P500 Index.

We may categorize a particular VaR predictive result as a binary variable, taking on the value of 1 when an exception arises, and 0 when not. Such a binary event, featuring only two potential outcomes, can be suitably modeled through the application of a bernoulli distribution¹².

In undertaking 1,006 VaR predictions for the S&P500 Index, we can apply the binomial distribution as an extrapolation of the bernoulli distribution, accounting for multiple independent VaR estimations. This distribution effectively models the cumulative count of exceptions arising from a predefined number of independent bernoulli VaR predictions.

By using a binomial distribution, we can measure the probability of the VaR level being exceeded on m or more days as:

$$\mathbf{Prob}_{(\# \text{ exceptions} \geq m)} = \sum_{k=m}^n \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \quad (18)$$

where:

n is the number of days use for the back-test (1,006 days)

m is the number of exceptions for a particular VaR model

p is the probability of the 95% VaR being exceeded on any given day (‘ideal’ VaR model: 5%)

¹²Olofsson, P., 2005. Probability, Statistics, and Stochastic Processes. Wiley-Interscience, pp. 129-134.

And the probability of the VaR level being exceeded on m or less days as:

$$\mathbf{Prob}_{(\# \text{ exceptions} \leq m)} = 1 - \mathbf{Prob}_{(\# \text{ exceptions} > m)} = \sum_{k=0}^m \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \quad (19)$$

When estimating the 5th percentile of the loss distribution, the likelihood of an exception on a given day is five percent. By using a binomial distribution to back-test the five models developed in this work, we consider two alternative hypothesis; the first is that the probability of an exception on any given day is $p = 5\%$, however the second alternative, varies based on the deficit ratio (m/n) observed for a particular VaR model. Specifically¹³:

1. If the observed deficit ratio exceeds p , the alternative hypothesis (H_1) posits that the true daily probability of an exception is greater than p .
2. Conversely, if the deficit ratio is less than p , the alternative hypothesis contends that the true daily probability of an exception is beneath p .

Given these two potential alternative hypotheses, the decision to accept or reject the model is contingent upon both the observed deficit ratio and a confidence level ($\alpha = 95\%$). If the ratio exceeds p , equation 18 is employed to calculate the likelihood that the VaR level will be surpassed on m or more days. On the other hand, when the deficit ratio is below p , equation 19 is used to determine the probability of the VaR level being exceeded on m or fewer days. In both scenarios, the model is accepted if the outcome exceeds the threshold of 5% (which is $1 - \alpha$).

Model	m	m/n	Confidence level = 95%		
			Equation Used	Equation Result	Back-Test Result
Model 1: EWMA($\lambda = 0.82$)	35	3.48%	equation 19	1.28%	Reject Model
Model 2: GARCH(1, 1)	27	2.68%	equation 19	0.02%	Reject Model
Model 3 v2: GJR-GARCH(1, 1)	38	3.78%	equation 19	3.97%	Reject Model
Model 4: EGARCH(1, 1)	68	6.76%	equation 18	0.58%	Reject Model
Model 5: Mix Model	52	5.17%	equation 18	36.78%	Accept Model

Table 1: Back-Testing VaR models using binomial distribution

Results using a binomial distribution approach to back-test the performance for each VaR model developed were shown in Table 1. At a five percent significance level, we should reject the first four models. Interestingly, the only model accepted using a five percent significance level, was the concluding one (Model 5), which synergistically combined both volatility models while taking into account the leverage effect.

A binomial distribution approach used for back-test a VaR model employed a one-tailed test. In the case of the initial three models evaluated in this study, as the deficit ratios were less than five percent, it was presumed that the probability of an exception was either exactly five percent or less than five percent. On the other hand, as the deficit ratios of the latter two volatility models (4 & 5) exceeding the 5% threshold, the assumption transitioned to a probability of five percent or more. However, in 1995 an economist of the Federal Reserve, Paul Kupiec¹⁴, introduced a new approach to back-test the VaR model performance by using a two-tailed test:

$$\mathbf{Kupiec}_{\text{Test}} = -2\ln[(1-p)^{n-m}p^m] + 2\ln\left[\left(1 - \frac{m}{n}\right)^{n-m} \left(\frac{m}{n}\right)^m\right] \quad (20)$$

¹³Hull, J.C., 2018. Back-Testing. In: Risk Management and Financial Institutions. 5th ed. Hoboken, New Jersey: John Wiley & Sons, Inc., pp.285-288.

¹⁴Kupiec, P., 1995. Techniques for Verifying the Accuracy of Risk Management Models. Journal of Derivatives, 3, pp. 73-84.

And the test-statistic follows a chi-square with one degree of freedom:

$$\text{Kupiec}_{\text{Test}} \sim \chi^2_{(df=1)}$$

In Kupiec's test, extreme values can be attributed to a substantially low or high count of exceptions. By definition a chi-square variable, with one degree of freedom, presents a five percent likelihood of exceeding a critical value of 3.84. Adopting this methodology to back-test the precision of a 95% VaR model should reject the model whenever the expression in equation 20 surpasses this critical value.

H_0 : Accept Model

H_1 : Reject Model

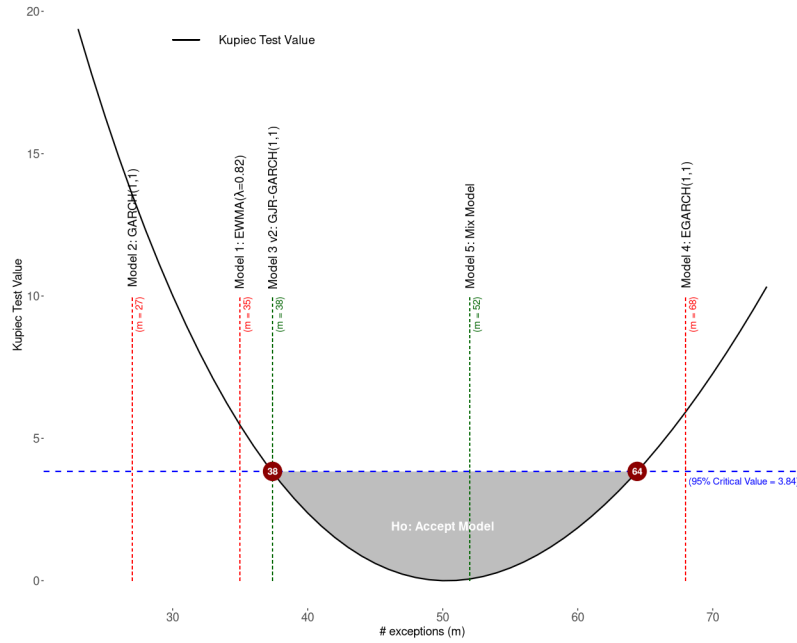


Figure 25: Kupiec test statistic as a function of the number of exceptions.

As illustrated in Figure 25, there exists a quadratic relationship between the Kupiec test value and the number of exceptions (black line). This indicates that extreme counts of exceptions, either low or high, correspond to higher Kupiec test values. In our analysis spanning 1,006 days of back-testing for each VaR model, the number of exceptions corresponding to a Kupiec test value of 3.84 (representing a 95% critical value) are 38 and 64. This implies that at a 95% confidence level, a VaR model's performance is acceptable if the number of exceptions lies between these two numbers. However, if the observed exceptions fall outside these bounds, the model's performance is considered unsatisfactory. Notably, Model 3 v2 underestimated the negative return of the S&P500 only 38 times, positioning it at the lower acceptance threshold. In contrast, and in agreement with our initial back-test results utilizing the binomial distribution, Model 5 showed the best results as the number of exceptions for this model was approximately centered within the acceptable range.

This particular model's efficiency is apparent in Figure 24, where it demonstrated 52 exceptions, translating into a deficit ratio of 5.17%. This ratio is inline with an 'ideal' model aimed at estimating the 5th percentile of the loss distribution. It is worth noting that these exceptions were uniformly dispersed over the five-year data period used for VaR estimation, further substantiating the model's effectiveness.

From the Kupiec back-test analysis, it becomes evident that Model 5 stands out as the most effective VaR model. Its efficacy stems from its dynamic integration of two volatility models that accommodate the leverage

effect, utilizing AIC score for model selection. These findings align with our first method of assessing the 95% VaR model performance, where compared the deficit ratio (m/n) against an ‘ideal’ 95% VaR model which should have a deficit ratio of five percent.

Conclusion

...at the end...

References

1. Black, F. (1976). *Studies of Stock Price Volatility Changes*. Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association, 177-181.
2. Bollerslev, T. (1986). *Generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics, 31(3), pp. 307-327.
3. Christoffersen, P. (2012). *Elements of Financial Risk Management*(2nd ed., pp. 65-143). Waltham, MA: Academic Press.
4. Fiorentini, G., Calzolari, G. & Panattoni, L., 1996. *Analytic Derivatives and the Computation of Garch Estimates*. Journal of Applied Econometrics.
5. Glosten, L.R., Jagannathan, R. and Runkle, D.E., 1993. *On the relation between the expected value and the volatility of the nominal excess return on stocks*. The Journal of Finance, 48(5), pp.1779-1801.
6. Hull, J.C., 2018. *Choice of Parameters for VaR and ES*. In: *Risk Management and Financial Institutions*. 5th ed. Hoboken, New Jersey: John Wiley & Sons, Inc., pp.278.
7. Hull, J.C., 2018. *The Exponentially Weighted Moving Average Model*. In: *Risk Management and Financial Institutions*. 5th ed. Hoboken, New Jersey: John Wiley & Sons, Inc., pp.225-228.
8. Hull, J.C., 2018. *Back-Testing*. In: *Risk Management and Financial Institutions*. 5th ed. Hoboken, New Jersey: John Wiley & Sons, Inc., pp.285-288.
9. J.P. Morgan/Reuters Financial, 1996. *RiskMetrics - Technical Document*. 4th ed. [pdf] New York: J.P. Morgan. Available at: Link [Accessed June 14th, 2023]
10. Kupiec, P., 1995. *Techniques for Verifying the Accuracy of Risk Management Models*. Journal of Derivatives, 3, pp. 73-84.
11. Lambert, R.A., 2012. *Financial Literacy for Managers*. University of Pennsylvania Press. (Chapter 'Your Company's Financial Health: What Financial Statements Can Tell You', p.33)
12. Nelson, D.B. (1991). *Conditional Heteroskedasticity in Asset Returns: A New Approach*. Econometrica, 59(2), 347-370.
13. McNeil, A. J. (2015). *Quantitative Risk Management: Concepts, Techniques, and Tools* (Revised edition).
14. Olofsson, P., 2005. *Probability, Statistics, and Stochastic Processes*. Wiley-Interscience, pp. 129-134.
15. Tsay, R.S. (2010) *3.8 The Exponential GARCH Model*. In: *Analysis of Financial Time Series*. 3rd edn. Hoboken, New Jersey: John Wiley & Sons, Inc., pp. 124-130.
16. Tsay, R.S. (2014). *Multivariate Time Series Analysis: With R and Financial Applications*. Hoboken, New Jersey: John Wiley & Sons Inc.