Backtesting Value-at-Risk: An Alternative Methodology*

Giuseppe Arbia[†] Rong Chen[‡] Giuseppe Corvasce[§] Ruey S. Tsay[¶]

Abstract

The paper proposes a novel statistical methodology for backtesting Value-at-Risk (VaR) models. The technique relies on the Ljung-Box test for the size of the hits, computed as the distance between the observed returns and the one step ahead forecasted VaR, when a violation occurs. The test determines whether or not the size of the hits are independent and identically distributed; whether or not the model shows lack of fit. The empirical analysis is applied to the S&P500 index, considering the levels of the VaR at 95% and 99%.

Keywords: Time-series models, Value-at-Risk, backtesting
JEL Classification: C22, C40, C52

*The authors would like to thank Yakov Amihud, Frank Diebold, Robert F. Engle, Victor Ng, Jacob Boudoukh, Robert Whitelaw, Matthew Richardson as well as several participants of the conferences and seminars organized by the GARP New York and Chicago Chapters, the University of Alberta, the Department of Finance at the University of Luxembourg, The Wharton School at the University of Pennsylvania, Università Cattolica del Sacro Cuore, the University of Chicago - Booth School of Business, the Allied Social Science Association, the Chicago Board Options Exchange, The Paris EUROPLACE International Financial Forum, The U.S. Department of the Treasury, the Office of Financial Stability (OFS), the Office of Financial Research (OFR), the International Organization of Securities Commissions (OICV-IOSCO), the Center for Financial Statistics and Risk Management at Rutgers University, its advisors and the staff members. The authors also thank several consultants, advisors, hedge/private equity and family funds as well as banks and insurance companies for comments received.

[†]Università Cattolica del Sacro Cuore - Department of Statistics; Address: Largo Francesco Vito 1, 00168 Rome, Italy. Email: giuseppe.arbia@unicatt.it

[‡]Rutgers University - Department of Statistics, School of Arts and Sciences; Address: Hill Center, 110 Frellinghuysen Road, Piscataway, NJ 08854-8019, USA. Email: rongchen@stat.rutgers.edu

[§]Rutgers University - Center for Financial Statistics and Risk Management; Address: Hill Center 110 Frelinghuysen Road Piscataway, NJ 08854.
University of Pennsylvania - Wharton School; Address: 3733 Spruce Street, Philadelphia, PA 19104.6340.
Harvard University - Massachusetts Hall, Cambridge, MA 02138 - USA
Society for Financial Studies; Mailing Address: PO Box 28494 Raleigh, NC 27611.
Email: giuseppecorvasce@gmail.com.

[¶]University of Chicago - Booth School of Business - Econometrics and Statistics. Address: 5807 S. Woodlawn Avenue, Chicago, IL 60637, U.S.A. Email: ruey.tsay@chicagobooth.edu

1. Introduction

The Value-at-Risk (VaR) which measures the quantile of the predicted distribution of gains and losses over a target horizon, constitutes the most popular measure of risk. Consequently, regulatory authorities need to put adequate ex-post techniques, validating or not the amount of risk taken by financial institutions. The standard assessment method of VaR consists in backtesting procedures. Jorion (2007) defines a backtesting procedure as a formal statistical framework that consists in verifying if actual trading losses are in line with projected losses. This procedure involves a comparison of model-generated VaR forecasts with actual returns and generally relies on testing over VaR violations (also called the Hits). A hit is said to occur when ex-post portfolio returns are lower than VaR forecasts.

Christoffersen (1998) points out that the puzzle of determining the accuracy of a VaR model can be reduced to the problem of determining whether the hit sequence satisfies two properties, namely the unconditional coverage property and the independence property. The hypothesis of unconditional coverage means that the expected frequency of observed violations is equal to $\alpha\%$. If the unconditional probability of violation is significantly higher than $\alpha\%$, it means that VaR model understates the portfolio's actual level of risk. The independence property means that if the model of VaR calculation is valid, then violations must be distributed independently. Campbell (2007) points out how the unconditional coverage property places a restriction on how often VaR violations may occur; whereas, the independence property restricts the ways in which these violations may occur.

In line with the conventional wisdom, this paper proposes a novel statistical procedure for backtesting VaR models. The technique relies on the Ljung-Box test for the size of the hits, computed as the distance between the observed returns and the one step ahead forecasted VaR, when a violation occurs. The test determines whether or not the size of the hits are independent and identically distributed; whether or not the autocorrelations for the size of the hits are non zero. The null hypothesis is the data are independently distributed. If the p-value of the test is greater than a certain threshold, it means that the model does not show lack of fit. The analysis is applied to the Standard & Poor's 500 index, considering the level of the VaR at 95% and 99%.

The paper is organized as follows. In section 2, it is reported an overview of the literature. Section 3 presents the statistical methodology for backtesting VaR models. Section 4 discusses the data and reports the descriptive statistics that supports the empirical analysis. In section 5, there is a discussion of the econometric models for one step ahead forecasting the VaR. Section 6 realizes an empirical application using the daily data of the S&P500 index. Finally, the last section concludes.

2. An Overview of Literature

Campbell (2007) proposes a survey of various tests on independence and unconditional coverage hypotheses for backtesting VaR models. Escanciano and Olmo (2010) show how the standard unconditional and independence backtesting procedures to assess VaR models in out-of-sample composite environments can be misleading. These tests do not consider the impact of estimation risk and therefore may use wrong critical values to assess market risk.

Another streamline of the literature uses the statistical properties of the duration between two consecutive hits. The idea is that if the one-period ahead VaR is correctly specified for a coverage rate α, the durations between two consecutive hits must have a geometric distribution with a success probability equal to $\alpha\%$. Christoffersen and Pelletier (2004) propose a test of independence. The idea behind the duration-based backtesting test consists in specifying a duration distribution that nests the geometric distribution and allows for duration dependence, so that the independence hypothesis can be tested by means of simple likelihood ratio (LR) tests. As pointed out by Haas (2007), this backtesting procedure sounds very interesting. It must be note that one have to specify a particular distribution under the alternative hypothesis. Pelletier and Wei (2016) propose the geometric-VaR test which utilizes the duration between the violations of VaR as well as the value of VaR.

Candelon et al. (2011) propose a new duration-based backtesting procedure for VaR forecasts. This analysis also relies on the GMM test framework proposed by Bontemps (2006) to test for the distributional assumption (i.e. the geometric distribution) and it is applied to the case of the VaR forecasts validity. The statistical technique tackles most of the drawbacks usually associated to duration based back-testing procedures. Lopez (1999) proposes the loss function evaluation method that is not a hypothesis-testing framework, but rather assigns to VaR estimates a numerical score that reflects specific regulatory concerns.

3. The methodology

As noted by the Basle Committee on Banking Supervision (1996), the magnitude as well as the number of exceptions, computed as the distance between the observed returns (r_t) and the one step ahead forecasted VaR, at the confidence level $0 \leq \alpha \leq 1$ (VaR_t^{α}) when a violation occurs, are a matter of regulatory concern.

This section proposes the Ljung-Box test (Ljung and Box 1978) for accepting or rejecting the VaR models, based on the size of the hits (S) at time t:

$$S_t = \begin{cases} r_t - VaR_t^{\alpha} & r_t \leq VaR_t^{\alpha} \\ 0 & r_t > VaR_t^{\alpha} \end{cases} . \tag{1}$$

The test determines whether or not the size of the hits are independent and identically distributed; whether or not the autocorrelations for the size of the hits are non zero. Therefore, the null hypothesis (H_0) is the data are independently distributed; with, the alternative hypothesis (H_a) being the data are not independently distributed. The null hypothesis also implies that the model does not show lack of fit.

The test statistic is:

$$Q = n\,(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}, \tag{2}$$

where n is the sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag k of the quantity S_t, and h is the number of lags being tested. Under the null hypothesis, the statistic Q asymptotically follows a $\chi^2_{(h)}$. For significance level δ, the critical region for rejection of the hypothesis of randomness is:

$$Q > \chi^2_{1-\delta,h} \tag{3}$$

where, $\chi^2_{1-\delta,h}$ is the $(1-\delta) - quantile$ of the chi-squared distribution with h degrees of freedom. We would like to **fail to reject** the null hypothesis. That is, we would like to see the p-value of the test be greater than a certain threshold, because this means the size of the hits are independent and identically distributed and the VaR model does not show lack of fit.

4. Data and descriptive statistics

The empirical analysis is applied to the close prices of the Standard & Poor's 500 ($S\&P500$) index, measured in US Dollars. The prices are transformed into returns by taking logarithmic differences of the closing daily prices. The analysis relies on daily data for the period from January 1st, 1928 through April 19th, 2023. Therefore, the sample covers the recent years of turmoil. The entire data period is divided into an estimation window (January 1st, 1928 to December 31, 2008) and a test window (January 1st, 2000 to April 19th, 2023). Thus, the empirical analysis works with 23938 observations and generates 5861 out-of-sample VaR forecasts, at 95% and 99%.

[Please Insert Table 1 around here]

Table 1 reports the descriptive statistics for the close prices of the S&P 500 index. From January 1st, 1928 to April 19th, 2023, the S&P 500 index respectively reports an average value of 580.735 USD and a median value of 101.260 USD, reaching a maximum value of 4796.560 USD on December 31st, 2021. In particular, during the estimation window, the unconditional mean for the S&P 500 is equal to 19.937 with a standard deviation of 1.896 and a level of skewness equals to 0.450; whereas, during the test window, the unconditional mean for the close prices of the S&P 500 index is equal to 1900.885, with a standard deviation of 988.958.

On October 19th, 1987, the S&P 500 registers its largest daily percentage loss, falling 20.47 percent. The one-day crash is known as "Black Monday". Despite the losses, the S&P 500 still closed positive for the year. The S&P 500 index reaches an all-time intraday high of 1552.87 USD, on March 24th, 2000, during the dot-com bubble. On October 9th, 2007, the index closes at a record high of 1565.15 USD, the highest close prior to the financial crisis of 2007–2008. Two days later, the index hits an intraday record high of 1576.09 USD. It did not regain this closing level until March 28th, 2013. On February 5th, 2018, after months of low volatility, the S&P 500 registers a new largest daily point loss of 113.19 points, equivalent to more than 4%. Three days later, the index suffered another heavy loss of nearly the same amount.

On October 13th, 2008, the S&P 500 marks its best daily percentage gain, rising 11.58 percent. It also registers its then-largest single-day point increase of 104.13 points. While on pace for the worst December performance since the Great Depression, the S&P 500 registers a new largest daily point gain of 116.60 points on December 26th, 2018, which translates to roughly 5% on the index. On February 19th, 2020, the S&P 500 index reached its highest point in the bull market that started from the low point on March 9th, 2009, closing at 3386.15 USD.

The S&P 500 index suffered its worst daily decline since 1987's Black Monday, falling 9.5 percent on March 16th, 2020, as a result of anxiety about the coronavirus pandemic. The decline of more than 20% since its peak, only 16 trading days earlier, signaled the start of a bear market closing at 2,480.64 USD. On August 18th, 2020, the S&P 500 index closed at a record high of 3389.78 USD amid the ongoing COVID-19 pandemic in the United States.

5. The econometric models

This section proposes the econometric models for one step ahead forecasting the VaR, at the significance levels of 95% and 99%. The natural logarithmic variations of the close prices, at time t can be computed in the following way:

$$\triangle p_t = \left(\frac{p_t - p_{t-1}}{p_{t-1}}\right) \simeq log\left(\frac{p_t}{p_{t-1}}\right) = c\,(1) + c\,(2)\,\triangle p_{t-1} + \varepsilon_t, \tag{4}$$

where, $c\,(1)$ is the coefficient of the mean equation that describes the evolution of the daily close prices, at time t; $c\,(2)$ is the autoregressive component of the mean equation and ε is the residual component at time t. This section proposes four different specifications of the conditional variance process (σ^2) at time t, with normal and student t distributions of the errors. Therefore, we have the following models: GARCH(1,1), GJR-GARCH(1,1), EGARCH(1,1) and PARCH(1,1).

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model was introduced by Bollerslev (1986) and relies on the following specification:

$$GARCH\,(1,1):\;\; \sigma_t^2 = c\,(3) + c\,(4)\cdot\varepsilon_{t-1}^2 + c\,(5)\cdot\sigma_{t-1}^2, \tag{5}$$

where, $c\,(3)$ is the long term component of the conditional variance; $c\,(4)$ depicts the influence of the squared residuals at time $t-1$ and $c\,(5)$ depicts the persistence of the conditional variance.

The second model is the GJR-GARCH by Glosten et al. (1993). The generalized specification for the conditional variance is given by:

$$GJR-GARCH\,(1,1):\ \sigma_t^2 = c\,(3) + c\,(4)\cdot\varepsilon_{t-1}^2 + c\,(5)\cdot\varepsilon_{t-1}^2\cdot(\varepsilon_{t-1}<0) + c\,(6)\cdot\sigma_{t-1}^2. \quad (6)$$

In this model, good news and bad news ($\varepsilon_{t-1} < 0$) have differential effects on the conditional variance. Good news has an impact of $c\,(4)$, while bad news has an impact of $c\,(4)+c\,(5)$. If $c\,(5) > 0$, bad news increases volatility, and we say that there is a *leverage effect*. If $c\,(5) \neq 0$, the news impact is asymmetric. The coefficient $c\,(6)$ depicts the persistence of the conditional variance.

The third model is the Exponential GARCH proposed by Nelson (1991). The specification for the conditional variance is:

$$EGARCH\,(1,1):\ log\left(\sigma_t^2\right) = c\,(3) + c\,(4)\cdot abs\left(\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}}\right) + c\,(5)\cdot\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + c\,(6)\cdot log\left(\sigma_{t-1}^2\right). \quad (7)$$

Note that the left-hand side is the logarithm of the conditional variance. This implies that the leverage effect is exponential, rather than quadratic and that forecasts of the conditional variance are guaranteed to be non-negative. The presence of leverage effects can be tested by the hypothesis that $c\,(5) < 0$. The impact is asymmetric if $c\,(5) \neq 0$. Since $log\,\sigma^2$ may be negative, there are no sign restrictions for the parameters. The coefficient $c\,(6)$ depicts the persistence of the conditional variance.

Taylor (1986) and Schwert (1989) introduced the standard deviation GARCH model, where the standard deviation is modeled rather than the variance. This model, along with several other models, is generalized in Ding et al. (1993), with the Power ARCH specification. In the Power ARCH model, the power parameter of the standard deviation can be estimated rather than imposed. Therefore, the following expression is derived.

$$PARCH\,(1,1): \left(\sqrt{\sigma_t^2}\right)^{c(7)} = c\,(3) + c\,(4) \cdot \left(abs\,(\varepsilon_{t-1}) - c\,(5) \cdot \varepsilon_{t-1}\right)^{c(7)} + c\,(6) \cdot \left(\sqrt{\sigma_{t-1}^2}\right)^{c(7)}, \quad (8)$$

where $c\,(7) > 0$, and the absolute value of $c\,(5)$ is smaller or equal than 1. The coefficient $c\,(6)$ depicts the persistence of the conditional variance.

6. Empirical Results

The estimation results rely on the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Roger 1987) that is an iterative method for solving unconstrained nonlinear optimization problems. It belongs to quasi-Newton methods and seeks a stationary point of a function, reachable when the gradient is zero. The optimization algorithm begins at an initial estimate for the optimal values and proceeds iteratively to get better estimates at each stage, till when there is a convergence for finding the solutions. For simplicity, the maximum number of iterations is fixed to n. 5,000 and the convergence rate to 1e-06. The step method is based on the Levenberg-Marquardt algorithm (Levenberg 1944; Marquardt 1963) that is more robust than the Gauss-Newton algorithm, since it allows to derive solutions even if the algorithm starts very far o from the final minimum. In cases with multiple minima, the algorithm converges to the global minimum only if the initial guess is already somewhat close to the final solution. The estimation procedure also accommodates the Huber-White estimator (Huber 1967; White 1980), that allows to derive the variance/covariance matrix considering the heteroscedasticity of the residuals.

[Please Insert Table 2 around here]

Table 2 contains the estimation results across the specifications of the conditional variance process: (i) GARCH(1,1) with normal errors; (ii) GARCH(1,1) with t-student errors; (iii) GJR-GARCH(1,1) with normal errors; (iv) GJR-GARCH(1,1) with t-student errors; (v) EGARCH(1,1) with normal errors; (vi) EGARCH(1,1) with t-student errors; (vii) PARCH(1,1) with normal errors; (viii) PARCH(1,1) with t-student errors. The mean equation follows an autoregressive process of order 1 with a constant. The estimated coefficients are related to the entire period from January 1st 1928 to April 19th, 2023.

All the coefficients of the models are statistically significant at the level of 1%. The estimated coefficient $c(2)$ that depicts the autoregressive component of the mean equation ranges from 0.06, for the AR(1)-EGARCH(1,1) with t-student errors, to 0.071 for the AR(1)-GJR-GARCH(1,1) with normal errors. Across specifications, the estimated coefficients that depict the persistence of the conditional variance range from 0.896, for the AR(1)-GARCH(1,1) with normal errors, to 0.988 for the AR(1)-EGARCH(1,1) with t-student errors.

[Please Insert Figure 1 and Figure 2 around here]

The coefficients that determine the conditional volatility are also estimated for the period between January 1st, 1928 and December 31th, 1928 and taken into consideration for the period from January 1st, 2000 to April 19th, 2023[1], with the aim to compute the one step ahead forecasted VaR. Figure 1 reports its evolution at the significance levels of 95% and 99%, based on the AR(1)-GARCH(1,1) process with normal (Figure 1.1) and t-student (Figure 1.2) errors. The mean and the median of the one step ahead forecasted VaR at 95% with normal and t-student errors are respectively equal to 0.020 and 0.018; whereas, the mean and the median of the one step ahead forecasted VaR at 99% with normal and t-student errors are respectively equal to 0.024 and 0.021.

The evolution of the one step ahead forecasted VaR at the significance levels of 95% and 99%, based on the AR(1)-GJR-GARCH(1,1) model, is reported in Figure 2. The mean and the median for the one step ahead forecasted VaR at 95% with normal and t-student errors

[1] The one-step-ahead forecast is a technique used in time series forecasting. It is used to evaluate how well a model would have done if you were forecasting for one day ahead

are respectively equal to 0.019 and 0.017; whereas, the mean and the median for the one step ahead forecasted VaR at 99%, with normal and t-student errors, are respectively equal to 0.022 and 0.020.

[Please Insert Figure 3 and Figure 4 around here]

The evolution for the one step ahead forecasted VaR at the significance levels of 95% and 99% is also respectively shown for the AR(1)-EGARCH(1,1) model (Figure 3) and the AR(1)-PARCH(1,1) process (Figure 4).

6.1 The Size Of The Hits And The Test Of Hypothesis

The size of the hits is computed as the distance between the observed returns of the S&P500 and the one step ahead forecasted Value at Risk (VaR) when a violation occurs (see equality n. 1), considering the levels of significance at 95% and 99%. Table 3 reports the descriptive statistics for the size of the hits across the models.

[Please Insert Table 3 around here]

Considering the level of confidence for the forecasted VaR at 95%, the AR(1)-GARCH(1,1) model reports a mean for the size of the hits equals to -0.020 and a standard deviation of 0.015 (for normal errors) and 0.014 (for t-student errors). The mean decreases to -0.019 for the rest of the models with a standard deviation equals to 0.013 for the AR(1)-GJR-GARCH(1,1) model and 0.012 for the AR(1)-EGARCH(1,1) and AR(1)-PARCH(1,1) models.

The descriptive statistics for the size of the hits are also reported for the level of confidence of the VaR at 99%. The mean of the size of the hits is equal to -0.024 with a standard deviation equals to 0.016 and 0.015 for the AR(1)-GARCH(1,1) model. The mean decreases to -0.023 for the AR(1)-EGARCH(1,1) and for the AR(1)-PARCH(1,1) models with a standard deviation equals to 0.013.

The size of the hits allows to evaluate the Ljung Box test. It determines whether or not the size of the hits are independent and identically distributed; whether or not the autocorrelations for the size of the hits are non zero.We would like to fail to reject the null hypothesis. That is, we would like to see the p-value of the test be greater than 0.05, because this means the size of the hits are independent and identically distributed and the VaR model does not show lack of fit.

[Please Insert Table 4 around here]

Table 4 reports the autocorrelation, the partial autocorrelation for the first ten lags of the Ljung Box test and the related p-values, across specifications of the conditional processes. The results show that the VaR models reject the null hypothesis, since the p-value of the test is smaller than 0.05, implying that the size of the hits are serially correlated and the VaR models constructed on the conditional volatility processes show lack of fit.

7. Conclusion

This paper proposes a novel methodology for backtesting Value-at-Risk (VaR) models, relying on the Ljung-Box test to evaluate the independence and identically distributed nature for the size of the hits. The empirical application, focused on the S&P 500 index, explores VaR models at 95% and 99% confidence levels across different specifications of conditional variance processes, including GARCH, GJR-GARCH, EGARCH, and PARCH models. The results indicate that, while the tested models show statistical significance across the various parameters, the size of the hits often exhibits serial correlation. This suggests that the VaR models, despite being robust in forecasting, demonstrate a lack of fit in capturing the complete distributional dynamics of returns. The persistence of this issue across the different model typologies emphasizes the need for improved statistical methodologies in the area of VaR backtesting.

The methodology proposed in this paper suggests an alternative and a rigorous framework for backtesting VaR models, offering a formal tool in order to identify limitations in their predictive accuracy. This approach could help both regulatory bodies and financial institutions to refine their risk assessment techniques, potentially leading to more reliable and resilient financial systems.

VaR models represent an evergreen area of research that is continuously demanding improvements and refinements. As a consequence, future research in the field could explore several directions to further enhance the robustness and applicability of the approach presented in the current paper. A first direction could refer to model enhancements and could involve the investigation of advanced machine learning or hybrid approaches that could better capture the non-linearities and asymmetries which are typical of financial time series data. A second direction of research could extend the proposed methodology to a broader range of applications, e. g. considering other asset classes or financial markets to evaluate its generalizability. A third area of research could consider the refinements in the dynamic backtesting frameworks aiming at developing adaptive strategies that account for the evolving market conditions, such as regime shifts or structural breaks. Finally, research could focus on regulatory aspects and could examine the possibility of integrating the proposed backtesting methodology into existing regulatory frameworks, thus ensuring compatibility with practical risk management needs. By addressing these further areas of research, future work could contribute to the development of more accurate and versatile risk management models, supporting both academic and industry advancements in financial econometrics.

Some references

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D., 1999. Coherent Measures of Risk. Mathematical Finance, 9, 203-228.

Basel Committee of Banking Supervision, 1996. Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements. Available at www.bis.org.

Berkowitz, J., Christoffersen, P., and Pelletier, D., 2011. Evaluating Value-at-Risk models with desk-level data. Management Science, 57, 2213-2227.

Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, 307–327.

Bontemps, C., 2006. Moment-based tests for discrete distributions, Working Paper.

Brockwell, P., and Davis, R., 2002. Introduction to Time Series and Forecasting (2nd ed.). Springer, 35–38.

Campbell, S. D., 2007. A review of backtesting and backtesting procedures. Journal of Risk, 9(2), pp 1-18.

Candelon, B., Colletaz, G., Hurlin, C., and Tokpavi, S., 2011. Backtesting Value-at-Risk: A GMM duration based test. Journal of Financial Econometrics, 9(2), 314-343.

Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review, 39, 841-86.

Christoffersen, P. F., 2003. Elements of Financial Risk Management. Boston: Academic Press.

Christoffersen, P. and Pelletier, D., 2004. Backtesting Value-at-Risk: A duration based approach. Journal of Financial Econometrics, 2, 84-108.

Ding, Z., Granger, C. W. J., and Engle, R. F., 1993. A Long Memory Property of Stock Market Returns and a New Model. Journal of Empirical Finance, 1, 83-106.

Dowd, K., 2002. Measuring Market Risk. John Wiley & Sons Ltd., New York.

Engle, R.F., 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. Econometrica, 50, 987-1008.

Engle, R.F., and Bollerslev, T., 1986. Modelling the Persistence of Conditional Variances. Econometric Reviews, 5(1), 1-50.

Engle, R.F., and Manganelli, S., 2004. CAViaR: Conditional Autoregressive Value-at-Risk by Regression Quantiles. Journal of Business and Economic Statistics, 22, 367-381.

Escanciano, J. C., and Olmo, J., 2010. Backtesting parametric Value-at-Risk with estimation risk. Journal of Business and Economic Statistics, 28, 36-51.

Escanciano, J. C., and Olmo, J., 2011. Robust backtesting tests for Value-at-risk models. Journal of Financial Econometrics, 9, 132-161.

Glosten, L., Jagannathan, R. and Runkle, D., 1993. On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. Journal of Finance, 48, 1779-1801.

Haas, M., 2006. Improved duration-based backtesting of value-at-risk. Journal of Risk, 8, 17-38.

Huber, P. J., 1967. The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 5, 221–233.

Jorion, P., 2001. Value at Risk, second edition. McGraw-Hill, New York.

Kupiec, P. H., 1995. Techniques for Verifying the Accuracy of Risk Measurement Models. Journal of Derivatives, 3, 73-84.

Ljung, G. M., Box, G. E. P., 1978. On a Measure of a Lack of Fit in Time Series Models. Biometrika, 65(2), 297–303.

Levenberg, K., 1944. A Method for the Solution of Certain Non-Linear Problems in Least Squares. Quarterly of Applied Mathematics, 2(2), 164-168.

Lopez, J. A., 1999. Methods for evaluating value-at-risk estimates. Economic Review: Federal Reserve Bank of San Francisco, 2, 3-17.

Lopez, J. A., 1999. Regulatory Evaluation of Value-at-Risk Models. Journal of Risk, 1, 37-64.

Marquardt, D., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. SIAM Journal on Applied Mathematics. 11(2), 431–441.

McNeil, A.J., Frey, R., and Embrechts, P., 2015. Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press, Princeton, New Jersey.

Nelson, D. B., 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach.

Econometrica, 59(2), 347-70.

Pelletier, D. and Wei, W., 2016. The geometric-VaR backtesting method. Journal of Financial Econometrics, 14(4), 725–745.

Perignon, C., and Smith, D., 2008. A new approach to comparing VaR estimation methods. Journal of Derivatives, 16, 54-66.

Roger, F., 1987. Practical Methods of Optimization. (2nd ed.) New York: John Wiley & Sons.

Schwert, G. W., 1989. Why does market volatility change over time? Journal of Finance, 44, 1115-1153.

Taylor, S., 1986. Modelling Financial Time Series, Wiley, Chichester

White, H., 1980. A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica, 48(4), 817–838.

Table 1.
Descriptive Statistics

The table contains the descriptive statistics (mean, median, max., min., std. dev. skewness and kurtosis) for the S&P 500 index related to the following periods: (i) January 1st 1928 to April 19th, 2023 (**entire period**); (ii) January 1st 1928 to December 31th, 1928 (**estimation window**); (iii) January 1st, 2000 to April 19th, 2023 (**test window**).

Statistics	Entire Period	Estimation Window	Test Window
Mean	580.735	19.937	1900.885
Median	101.260	19.545	1439.030
Max.	4796.560	24.350	4796.560
Min.	4.400	16.950	676.530
Std. Dev	921.660	1.896	988.958
Skewness	2.188	0.450	1.207
Kurtosis	7.780	2.330	3.440

Table 2.
Estimation Results

The table contains the estimation results across the specifications of the conditional variance process: (i) GARCH(1,1) with normal errors; (ii) GARCH(1,1) with t-student errors; (iii) GJR-GARCH (1,1) with normal errors; (iv) GJR-GARCH(1,1) with t-student errors; (v) EGARCH(1,1) with normal errors; (vi) EGARCH(1,1) with t-student errors; (vii) PARCH(1,1) with normal errors; (viii) PARCH(1,1) with t-student errors. The mean equation follows an autoregressive process of order 1 with a constant. The estimated coefficients are related to the entire period from January 1st 1928 to April 19th, 2023. *, **, *** indicate the statistical significances at 10%, 5% and 1%.

Coefficients	**Model**							
	AR(1)-GARCH(1,1) normal errors	**AR(1)-GARCH(1,1) t-student errors**	**AR(1)-GJR-GARCH(1,1) normal errors**	**AR(1)-GJR-GARCH(1,1) t-student errors**	**AR(1)-EGARCH(1,1) normal errors**	**AR(1)-EGARCH(1,1) t-student errors**	**AR(1)-PARCH(1,1) normal errors**	**AR(1)-PARCH(1,1) t-student errors**
C(1) x 1000	0.494***	0.605***	0.276***	0.464***	0.249***	0.430***	0.254***	0.426***
C(2)	0.065***	0.057***	0.071***	0.063***	0.067***	0.060***	0.069***	0.061***
C(3) x 1000	0.001***	0.001***	0.001***	0.001***	-0.274***	-0.233***	0.027***	0.066***
C(4)	0.100***	0.089***	0.039***	0.031***	0.173***	0.154***	0.092***	0.085***
C(5)	0.896***	0.908***	0.096***	0.109***	-0.076***	-0.084***	0.405***	0.558***
C(6)			0.905***	0.907***	0.985***	0.988***	0.912***	0.921***
C(7)							1.355***	1.127***
t		5.753***		6.070***		6.067***		6.060***

Table 3.
Descriptive statistics for the size of the hits

The table reports the descriptive statistics (mean, median, max., min., std. dev., skewness and kurtosis) for the size of the hits, computed as the distance between the observed returns of the S&P500 and the one step ahead forecasted Value at Risk (VaR) when a violation occurs, considering the levels of the Value at Risk at 95% (**Panel 3.1**) and 99% (**Panel 3.2**). The descriptive statistics consider the **test window** from January 1st, 2000 to April 19th, 2023.

Panel 3.1: Descriptive statistics that consider the **VaR at 95%**

Size of the hits								
Statistics	**AR(1)-GARCH(1,1)**		**AR(1)-GJR-GARCH(1,1)**		**AR(1)-EGARCH(1,1)**		**AR(1)-PARCH(1,1)**	
	Normal	**t student**	**Normal**	**t student**	**Normal**	**t student**	**Normal**	**t student**
Mean	-0.020	-0.020	-0.019	-0.019	-0.019	-0.019	-0.019	-0.019
Median	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017	-0.017
Max.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Min.	-0.218	-0.219	-0.184	-0.185	-0.159	-0.163	-0.172	-0.172
Std. Dev.	0.015	0.014	0.013	0.013	0.012	0.012	0.012	0.012
Skewness	-3.114	-3.071	-2.558	-2.556	-2.175	-2.157	-2.258	-2.260
Kurtosis	23.058	24.178	19.128	19.211	14.950	14.928	16.004	16.137

Panel 3.2: Descriptive statistics that consider the **VaR at 99%**

Size of the hits								
Statistics	**AR(1)-GARCH(1,1)**		**AR(1)-GJR-GARCH(1,1)**		**AR(1)-EGARCH(1,1)**		**AR(1)-PARCH(1,1)**	
	Normal	**t student**	**Normal**	**t student**	**Normal**	**t student**	**Normal**	**t student**
Mean	-0.024	-0.024	-0.022	-0.022	-0.023	-0.023	-0.023	-0.023
Median	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020
Max.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Min.	-0.235	-0.236	-0.194	-0.196	-0.165	-0.169	-0.180	-0.180
Std. Dev.	0.016	0.015	0.014	0.013	0.013	0.013	0.013	0.013
Skewness	-3.225	-3.186	-2.571	-2.572	-2.160	-2.100	-2.214	-2.218
Kurtosis	23.837	25.290	19.463	19.595	15.197	14.624	15.817	15.999

Table 4.
Ljung-Box test

The table reports the Ljung Box test for the size of the hits, computed as the distance between the observed returns of the S&P500 and the one step ahead forecasted Value at Risk (VaR) when a violation occurs, considering the levels of the VaR at 95% and 99%. The test determines whether or not the size of the hits are independent and identically distributed; whether or not the autocorrelations for the size of the hits are non zero.We would like **to fail to reject** the null hypothesis. That is, we would like to see the p-value of the test be greater than 0.05, because this means the size of the hits are independent and identically distributed and the VaR model does not show lack of fit. The table reports the autocorrelation, the partial autocorrelation for the first ten lags of the Ljung Box test and the related p-values.

Panel 4.1: Model AR(1)-GARCH(1,1)

	Normal								t-student							
Lags	**Autocorrelation**		**Partial Autocorrelation**		**Q-Stat**		**p-value**		**Autocorrelation**		**Partial Autocorrelation**		**Q Stat**		**p-value**	
	95%	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**
1	0.296	0.366	0.296	0.366	514.32	785.58	0.000	0.000	0.245	0.304	0.245	0.304	352.16	540.55	0.000	0.000
2	0.363	0.425	0.302	0.336	1288.00	1843.90	0.000	0.000	0.311	0.362	0.267	0.297	921.01	1308.00	0.000	0.000
3	0.371	0.434	0.249	0.270	2096.50	2946.30	0.000	0.000	0.318	0.368	0.225	0.242	1514.10	2100.40	0.000	0.000
4	0.348	0.409	0.173	0.182	2806.20	3928.70	0.000	0.000	0.288	0.337	0.150	0.157	2000.10	2764.90	0.000	0.000
5	0.349	0.409	0.142	0.148	3519.10	4911.70	0.000	0.000	0.284	0.333	0.120	0.124	2474.10	3414.00	0.000	0.000
6	0.330	0.391	0.099	0.101	4157.90	5809.10	0.000	0.000	0.264	0.311	0.082	0.082	2884.00	3981.40	0.000	0.000
7	0.371	0.426	0.144	0.143	4964.50	6874.60	0.000	0.000	0.302	0.346	0.123	0.124	3419.30	4683.50	0.000	0.000
8	0.316	0.375	0.061	0.057	5552.00	7701.50	0.000	0.000	0.242	0.286	0.041	0.037	3762.60	5164.90	0.000	0.000
9	0.362	0.416	0.111	0.108	6320.80	8717.00	0.000	0.000	0.294	0.335	0.100	0.099	4269.80	5825.60	0.000	0.000
10	0.328	0.381	0.061	0.051	6953.90	9568.50	0.000	0.000	0.256	0.296	0.050	0.043	4655.20	6338.70	0.000	0.000

Panel 4.2: Model AR(1)- GJR-GARCH(1,1)

	Normal								t-student							
Lags	**Autocorrelation**		**Partial Autocorrelation**		**Q-Stat**		**p-value**		**Autocorrelation**		**Partial Autocorrelation**		**Q Stat**		**p-value**	
	95%	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**	**95%**	**99%**
1	0.344	0.412	0.344	0.412	693.88	994.64	0.000	0.000	0.321	0.387	0.321	0.387	605.70	878.51	0.000	0.000
2	0.267	0.301	0.169	0.158	1112.00	1525.80	0.000	0.000	0.261	0.295	0.176	0.171	1006.00	1388.90	0.000	0.000
3	0.204	0.227	0.081	0.071	1356.40	1828.60	0.000	0.000	0.203	0.227	0.089	0.080	1248.00	1690.60	0.000	0.000
4	0.152	0.168	0.035	0.028	1492.60	1994.80	0.000	0.000	0.152	0.169	0.039	0.031	1384.40	1858.40	0.000	0.000
5	0.116	0.127	0.018	0.013	1571.30	2088.70	0.000	0.000	0.116	0.127	0.018	0.013	1463.00	1953.00	0.000	0.000
6	0.105	0.113	0.028	0.028	1635.60	2163.40	0.000	0.000	0.106	0.115	0.029	0.029	1528.60	2030.30	0.000	0.000
7	0.124	0.129	0.061	0.060	1725.10	2261.50	0.000	0.000	0.126	0.132	0.062	0.062	1621.50	2133.20	0.000	0.000
8	0.100	0.106	0.020	0.015	1783.60	2327.10	0.000	0.000	0.100	0.106	0.021	0.016	1679.70	2199.40	0.000	0.000
9	0.127	0.133	0.060	0.062	1878.40	2430.80	0.000	0.000	0.129	0.136	0.062	0.064	1777.90	2307.80	0.000	0.000
10	0.118	0.124	0.035	0.030	1960.80	2521.80	0.000	0.000	0.119	0.125	0.036	0.031	1861.10	2400.10	0.000	0.000

Panel 4.3: Model AR(1)-EGARCH(1,1)

Lags	Normal								t-student							
	Autocorrelation		Partial Autocorrelation		Q-Stat		p-value		Autocorrelation		Partial Autocorrelation		Q Stat		p-value	
	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
1	0.272	0.321	0.272	0.321	434.17	605.78	0.000	0.000	0.235	0.281	0.235	0.281	325.14	463.04	0.000	0.000
2	0.197	0.216	0.132	0.126	660.89	879.08	0.000	0.000	0.196	0.217	0.149	0.150	550.73	739.07	0.000	0.000
3	0.149	0.160	0.073	0.067	790.93	1029.90	0.000	0.000	0.158	0.172	0.090	0.087	696.43	912.34	0.000	0.000
4	0.110	0.116	0.037	0.032	861.60	1109.00	0.000	0.000	0.118	0.128	0.046	0.041	778.78	1008.00	0.000	0.000
5	0.085	0.087	0.023	0.018	903.57	1153.30	0.000	0.000	0.095	0.100	0.029	0.025	831.71	1067.00	0.000	0.000
6	0.079	0.083	0.029	0.030	939.90	1193.40	0.000	0.000	0.086	0.093	0.030	0.031	875.41	1117.40	0.000	0.000
7	0.116	0.121	0.075	0.078	1018.90	1279.70	0.000	0.000	0.117	0.124	0.070	0.072	956.04	1207.30	0.000	0.000
8	0.084	0.088	0.020	0.015	1059.90	1325.60	0.000	0.000	0.082	0.087	0.020	0.016	995.58	1252.20	0.000	0.000
9	0.116	0.121	0.063	0.066	1138.30	1411.80	0.000	0.000	0.120	0.126	0.066	0.069	1079.50	1346.10	0.000	0.000
10	0.105	0.110	0.038	0.035	1203.70	1483.40	0.000	0.000	0.106	0.111	0.040	0.036	1145.50	1419.10	0.000	0.000

Panel 4.4: Model AR(1)-PARCH(1,1)

Lags	Normal								t-student							
	Autocorrelation		Partial Autocorrelation		Q-Stat		p-value		Autocorrelation		Partial Autocorrelation		Q Stat		p-value	
	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
1	0.243	0.289	0.243	0.289	346.20	489.48	0.000	0.000	0.224	0.267	0.224	0.267	293.97	417.18	0.000	0.000
2	0.211	0.236	0.162	0.167	608.05	816.35	0.000	0.000	0.205	0.229	0.163	0.170	541.25	724.67	0.000	0.000
3	0.179	0.198	0.106	0.105	796.80	1047.30	0.000	0.000	0.175	0.194	0.109	0.108	721.85	944.98	0.000	0.000
4	0.134	0.147	0.050	0.044	902.01	1173.40	0.000	0.000	0.131	0.143	0.053	0.047	823.29	1065.80	0.000	0.000
5	0.112	0.122	0.034	0.032	975.16	1260.50	0.000	0.000	0.110	0.119	0.036	0.032	894.32	1149.30	0.000	0.000
6	0.100	0.108	0.031	0.030	1033.40	1328.90	0.000	0.000	0.097	0.105	0.031	0.029	949.87	1214.20	0.000	0.000
7	0.123	0.132	0.065	0.067	1122.70	1430.80	0.000	0.000	0.122	0.129	0.065	0.066	1036.80	1312.50	0.000	0.000
8	0.088	0.095	0.018	0.014	1168.10	1484.00	0.000	0.000	0.085	0.091	0.018	0.014	1079.10	1361.60	0.000	0.000
9	0.130	0.140	0.071	0.076	1267.90	1598.90	0.000	0.000	0.127	0.136	0.070	0.074	1174.50	1470.80	0.000	0.000
10	0.113	0.119	0.038	0.033	1342.50	1681.90	0.000	0.000	0.109	0.115	0.039	0.034	1244.80	1548.40	0.000	0.000

Figure 1.
Value at Risk Estimation using the AR(1)-GARCH(1,1)

The figure reports the one step ahead forecasted Value at Risk (VaR) based on the **AR(1)-GARCH(1,1)** process with Normal (**Figure 1.1**) and t-student (**Figure 1.2**) errors. The bottom axis reports the date; whereas, the left axis reports the level of the VaR. The figures are related to the test window from January 1st, 2000 to April 19th, 2023.

Figure 1.1
VaR Estimation Using the AR(1)-GARCH(1,1) with Normal Errors

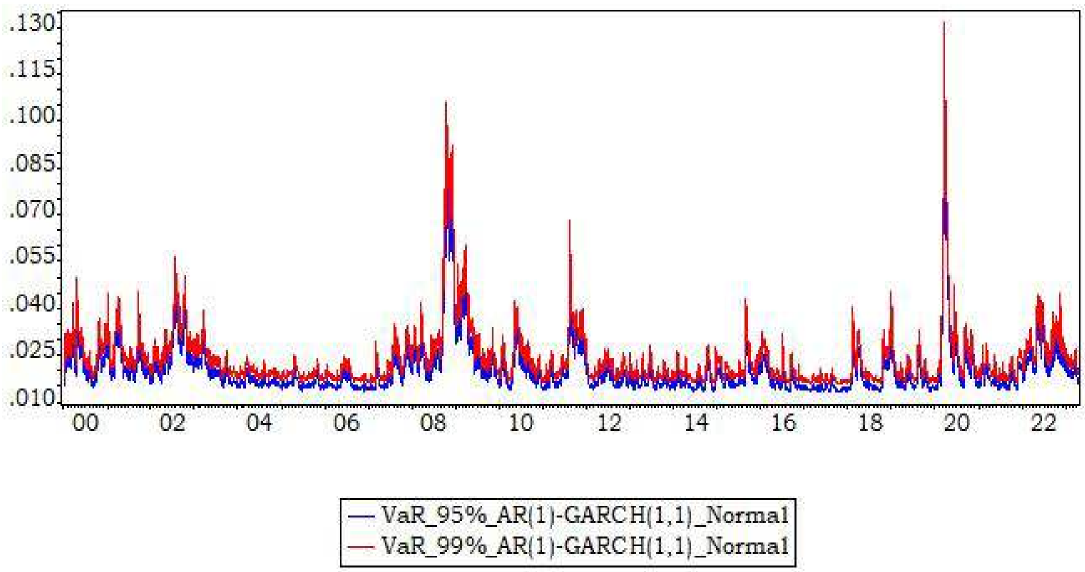

Figure 1.2
VaR Estimation Using the AR(1)-GARCH(1,1) with t-student Errors

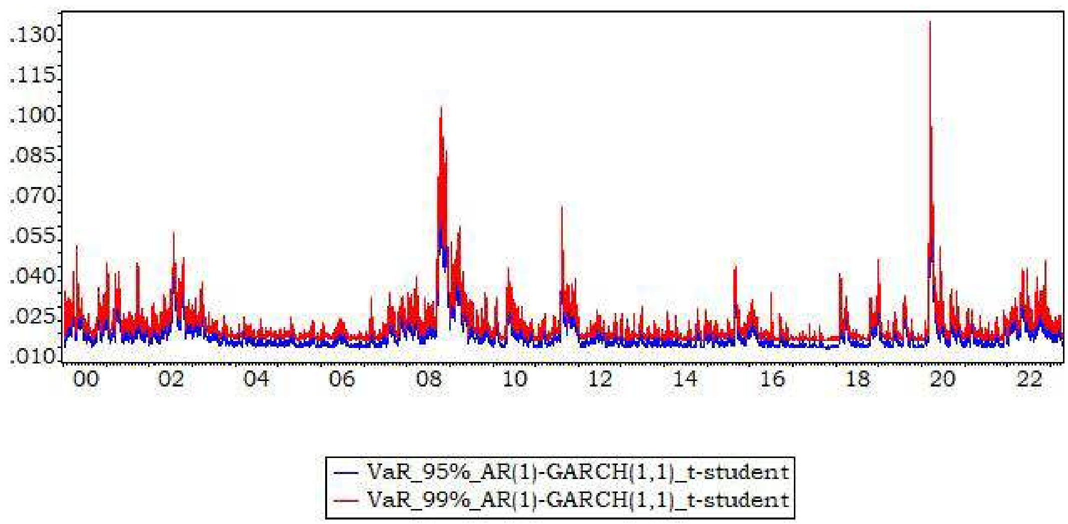

Figure 2.
Value at Risk Estimation using the AR(1)-GJR-GARCH(1,1)

The figure reports the one step ahead forecasted Value at Risk (VaR) based on the **AR(1)-GJR-GARCH(1,1)** process with Normal (**Figure 2.1**) and t-student (**Figure 2.2**) errors. The bottom axis reports the date; whereas, the left axis reports the level of the VaR. The figures are related to the test window from January 1st, 2000 to April 19th, 2023.

Figure 2.1
VaR Estimation Using the AR(1)-GJR-GARCH(1,1) with Normal Errors

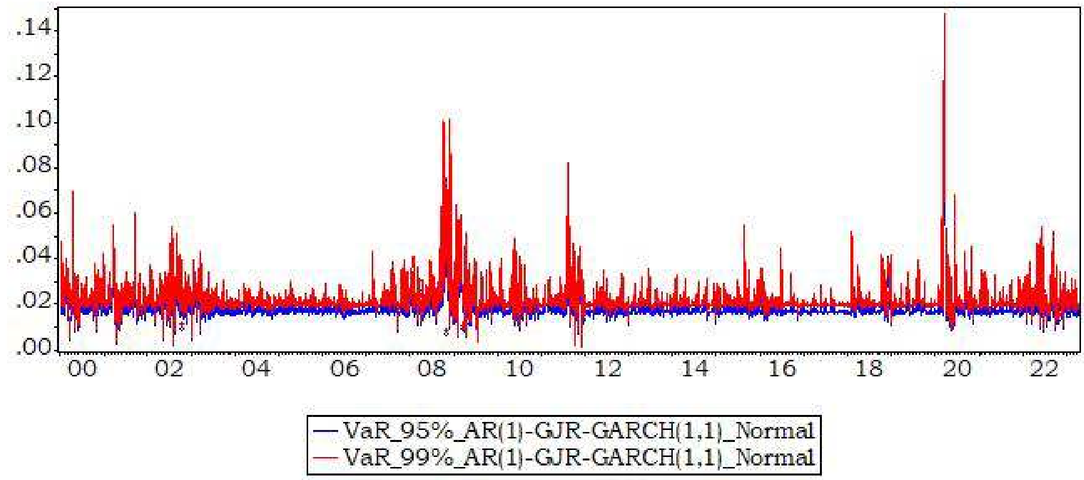

Figure 2.2
VaR Estimation Using the AR(1)-GJR-GARCH(1,1) with t-student Errors

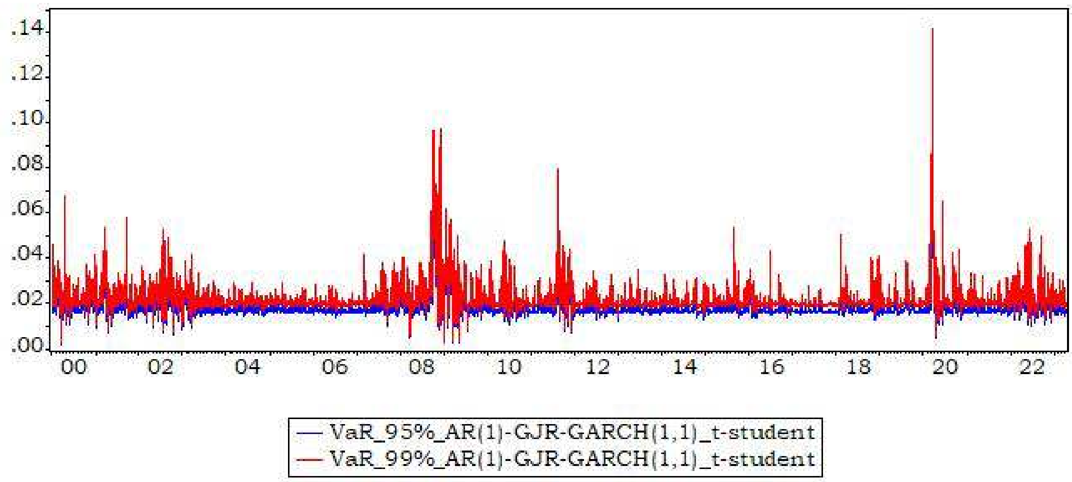

Figure 3.

Value at Risk Estimation using the AR(1)-EGARCH(1,1)

The figure reports the one step ahead forecasted Value at Risk (VaR) based on the **AR(1)-EGARCH(1,1)** process with Normal (**Figure 3.1**) and t-student (**Figure 3.2**) errors. The bottom axis reports the date; whereas, the left axis reports the level of the VaR. The figures are related to the test window from January 1st, 2000 to April 19th, 2023.

Figure 3.1

VaR Estimation Using the AR(1)-EGARCH(1,1) with Normal Errors

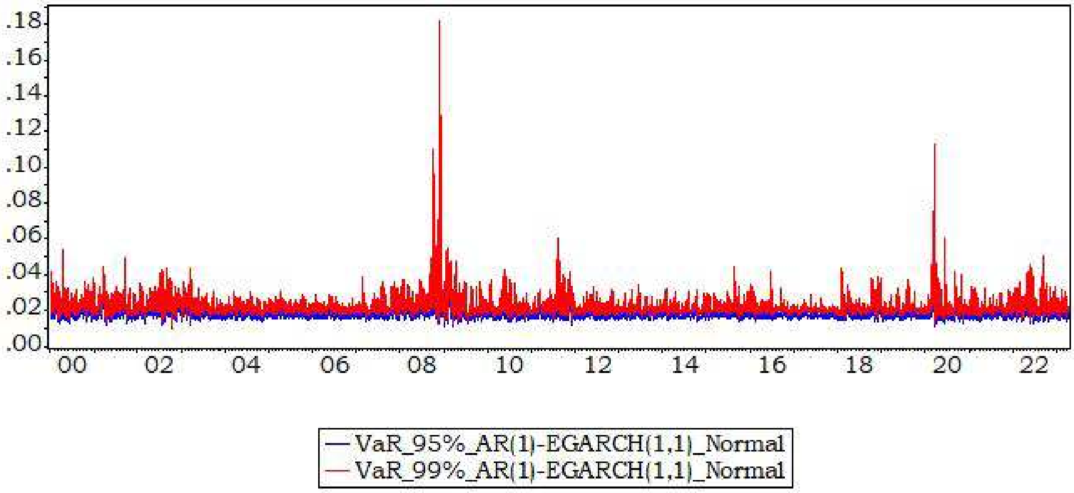

Figure 3.2

VaR Estimation Using the AR(1)-EGARCH(1,1) with t-student Errors

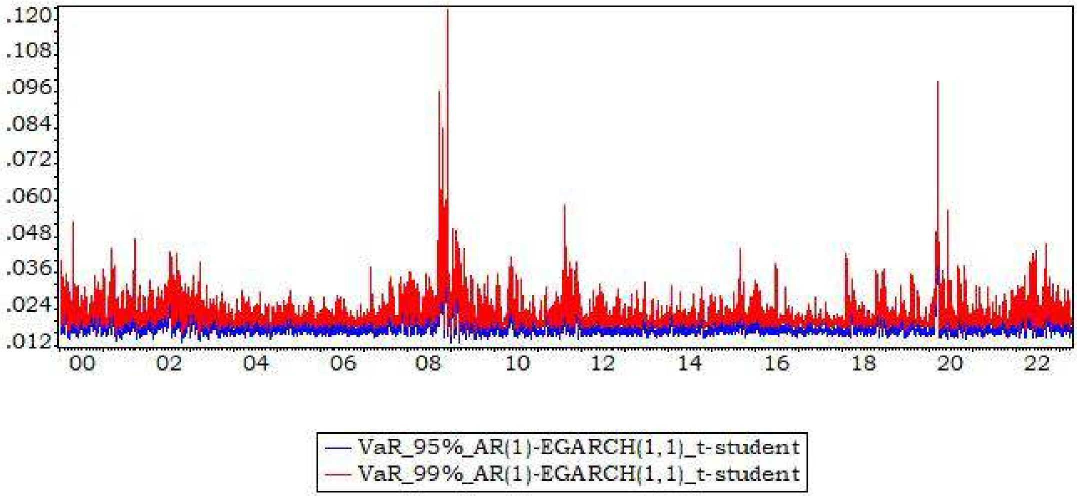

Figure 4.
Value at Risk Estimation using the AR(1)-PARCH(1,1)

The figure reports the one step ahead forecasted Value at Risk (VaR) based on the **AR(1)-PARCH(1,1)** process with Normal (**Figure 4.1**) and t-student (**Figure 4.2**) errors. The bottom axis reports the date; whereas, the left axis reports the level of the VaR. The figures are related to the test window from January 1st, 2000 to April 19th, 2023.

Figure 4.1
VaR Estimation Using the AR(1)-PARCH(1,1) with Normal Errors

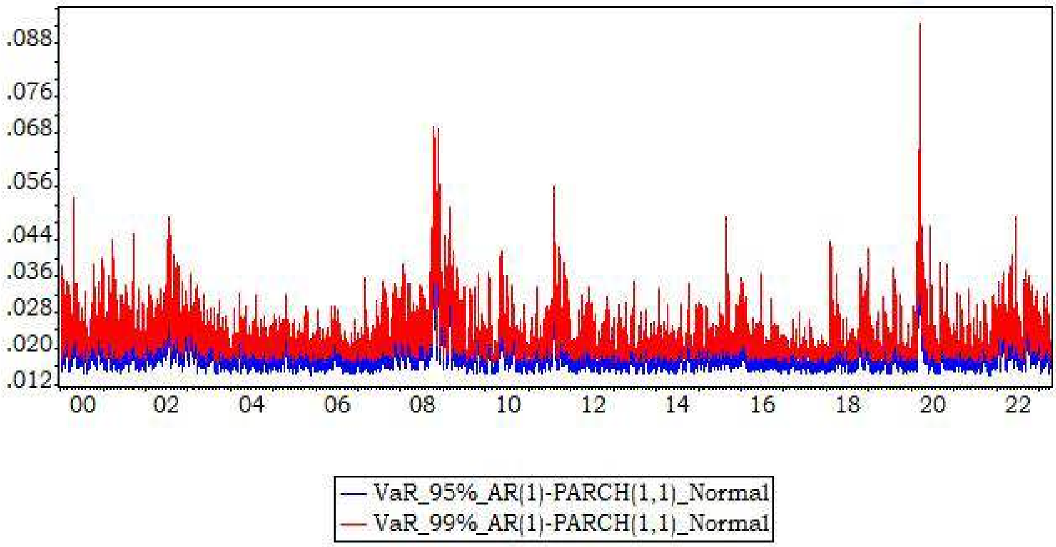

Figure 4.2
VaR Estimation Using the AR(1)-PARCH(1,1) with t-student Errors