

# Distribution Analysis of S&P 500 Financial Turbulence

Hugo Gobato Souto

BASF Nederland B.V., Arnhem, Netherlands

Email: hugo.gobato-souto@basf.com

**How to cite this paper:** Souto, H.G. (2023) Distribution Analysis of S&P 500 Financial Turbulence. *Journal of Mathematical Finance*, 13, 67-88.

<https://doi.org/10.4236/jmf.2023.131005>

**Received:** January 24, 2023

**Accepted:** February 18, 2023

**Published:** February 21, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

In 2010 a new financial risk measure was discovered, namely Financial Turbulence. Although it has been studied by other papers, its statistical distribution study is still missing. Knowing a financial phenomenon distribution is of importance when performing risk and portfolio management, especially when estimating parametric Value-at-Risk with Copulas and performing Monte Carlo simulations. Therefore, this paper explores the S&P 500 Financial Turbulence to determine its best distribution fit by making use of various residual measures and goodness-of-fit tests. Additionally, it makes use of in-sampling and out-sampling in the period between 01/11/2012 and 01/11/2022. The results of this research indicate that the Generalised Hyperbolic distribution describes the S&P 500 Financial Turbulence the best.

## Keywords

Financial Turbulence, Generalised Hyperbolic Distribution, S&P 500, Goodness-of-Fit Tests

---

## 1. Introduction

Understanding stock price fluctuations and their causes has always been a topic of interest in the field of finance. With the development of mathematical and statistical finance, the interest in this topic has only grown. Nowadays, it would be unimaginable for a financial organisation to invest without making use of the numerous existing portfolio and risk models. However, the use of inaccurate models can lead portfolio managers to overstate their portfolio robustness against risk [1]. This overstatement can result in unexpected losses and even social disbelief in financial mathematical techniques. A perfect example of such disbelief is the book “A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing” by Burton G. Malkiel [2], where a clear and

fair critique of the use of mathematical financial models by financial institutions can be found.

Presumably, the most famous risk parameter for financial mathematical models is stock volatility. It is a crucial input for many models related to portfolio management, risk mitigation and financial assets pricing, such as options. Stock volatility is so important that a large part of the existing financial literature was dedicated to examine its predictability using both univariate and multivariate forecasting models utilising a range of financial and economic indicators [3]-[10].

However, there exist two risk parameters that still require further exploration by the scientific community. These are Financial Turbulence (FT) and Absorption Ratio (AR). The first was devised by Kritzman & Li [11], and the second by Kritzman, Li, Page & Rigobon [12]. FT was derived from Mahalanobis distance and is given as:

$$d_t = (y_t - \mu)\Sigma^{-1}(y_t - \mu)' \quad (1)$$

where,

$d_t$  = turbulence for a particular time period t (scalar)

$y_t$  = vector of asset returns for period t ( $1 \times n$  vector)

$\mu$  = sample average vector of historical returns ( $1 \times n$  vector)

$\Sigma$  = sample covariance matrix of historical returns ( $n \times n$  matrix)

AR, on the other hand, makes use of the stocks covariance matrix eigenvectors in order to estimate their systemic risk, and its formula is given as:

$$AR = \frac{\sum_{i=1}^n \sigma_{E_i}^2}{\sum_{j=1}^N \sigma_{A_j}^2} \quad (2)$$

where,

AR = Absorption Ratio

$\sigma_{E_i}^2$  = variance of the  $i$ -th eigen vector, sometimes called eigenportfolio

$\sigma_{A_j}^2$  = variance of the  $j$ -th asset

$n$  = number of eigenvectors used to calculate AR

$N$  = number of assets

Salisu, Demirer & Gupta [13] showed that the use of such financial indicators can indeed improve out-of-sample predictive performance of stock market volatility models over both the short and long horizons. Their use also extends to portfolio management [14] [15].

However, there is currently no distribution analysis of FT and AR, unlike many other financial phenomena [16]-[21]. Even though one can make use of the Central Limit Theorem (CLT) in their analysis and portfolio management and/or risk management to compensate this lack of knowledge, there are many occasions in the financial field in which it is not possible to use this theorem, for instance when performing Value-at-Risk (VaR) analysis, Copulas, and Monte Carlo simulations. In addition, the misuse of the CLT can lead to problems, such as the underestimation of risk and extreme events possibility [22] [23] [24].

Needless to say, such a misuse can result in social disbelief in financial mathematical techniques.

Therefore, this study was performed to close this knowledge gap and improve the current financial models used in research and by financial institutions. Nonetheless, due to time and resource constraints, this paper author performed a thorough distribution analysis of only the FT of the famous S&P 500 index. Yet, the author strongly encourages the scientific community to do the same with AR, as well as with other indexes and markets.

This paper is structured as the following: Data and Methodology, Results, Conclusion, Bibliographical references. In the chapter “Data and Methodology”, the data selection procedure, in-sample and out-sample division and methodology used to perform the distribution analysis of the S&P 500 FT can be found. The chapter “Results” shows and discusses the results of the S&P 500 FT distribution analysis. Additionally, in the chapter “Conclusion, one can find the conclusion the S&P 500 FT distribution analysis results and their implications. Lastly, the chapter “Bibliographical references” lists the sources used in this research.

## 2. Data and Methodology

The data set used in this research was retrieved from Yahoo Finance through the use of the Python library `yfinance`. The time horizon was 5 years for both the in-sample and out-sample data. The in-sample is from 01/11/2017 until 01/11/2022, and the out-sample is from 01/11/2012 until 01/11/2017. Although having the out-sample set be further in the past than the in-sample set appears to be an unusual choice, it was made deliberately, in order to mitigate bias in the research. This is the case since the author has a better memory of the S&P 500 price and volatility development in the past 5 years than in the out-sample period. However, in theory the choice of having the in-sample period before or after the out-sample period ought to have no impact on the results of a distribution analysis since the goal of such an analysis is to find the distribution that best describes the random variable (in this case the S&P 500 FT) in any time period. That is, the results of the distribution analysis should not depend on time, otherwise its results are meaningless since one cannot be sure about the accuracy of their model results if they use the “best distribution” retrieved from the distribution analysis for a model that predicts future outcomes.

Due to the long time horizons used in the in-sample and out-sample, a few stocks that were present in the S&P 500 on 01/11/2022 do not have the historical data for the whole time-frame of both the in-sample and out-sample data. For the in-sample data, the stocks that did not have data throughout the whole time frame are 1) CARR, 2) CDAY, 3) CEG, 4) CTVA, 5) DOW, 6) FOX, 7) FOXA, 8) MRNA, 9) OGN, 10) OTIS, 11) VICI. Together they represent 2.19% of the total number of S&P 500 stocks and only 1.32% of the total S&P 500 market cap. Therefore, it can be concluded that their absence in the calculations did not have a significant effect on the final results of this research; and thus they were ex-

cluded from the calculations to make the calculations more coherent given the need for the covariance matrix in the FT equation (Equation (1)). For the out-sample data, the stocks that did not have data throughout the whole time frame are 1) CARR, 2) CDAY, 3) CEG, 4) CTVA, 5) DOW, 6) FOX, 7) FOXA, 8) MRNA, 9) OGN, 10) OTIS, 11) VICI, 12) ABBV, 13) ALLE, 14) ANET, 15) CFG, 16) CCTL, 17) PAYC, 18) PYPL, 19) QRVO, 20) CDW, 21) SEDG, 22) SYF, 23) WRK, 24) ZTS, 25) CZR, 26) ETSY, 27) FTV, 28) HLT, 29) HPE, 30) HWM, 31) INVH, 32) IQV, 33) IR, 34) KEYS, 35) KHC, 36) LW, 37) NWS, 38) NWSA, 39) NCLH. Together they represent 7.95% of the total number of S&P 500 stocks and 3.99% of the total S&P 500 market cap. Again, it can be concluded that their absence in the calculations did not have a significant effect in the final results of this research; and thus they were excluded from the calculations to make them more coherent given the need of the covariance matrix in the FT equation (Equation (1)).

The methodology below was used for the in-sample and out-sample data. The in-sample data was used to determine the best distribution(s), whereas the out-sample data was utilised to test the results. For the calculation of the S&P 500 FT, Equation (1) was used. Moreover, the programming language Python was used to create the algorithm for this calculation. This algorithm, as well as all calculations used for this research, can be found *here*.

After the S&P 500 FT calculation, the Python library Fitter was used to analyse the goodness-of-fit of the empirical data with 109 different distributions (those can be found in Appendix 1). This library estimates each distribution's parameters through the use of Log Maximum Likelihood Estimation (Log MLE).

Furthermore, to determine the goodness-of-fit of each distribution, this Python library made use of residual sum of squares (RSS), Equation (3), Akaike information criterion (AIC), Equation (4), and Bayesian information criterion (BIC), Equation (5). It used the following criterion: the lower the first measure and the higher the absolute value of the other two measures, the better goodness-of-fit a certain distribution has.

$$\text{RSS} = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (3)$$

where,

$y_i$  = empirical data point

$f(x_i)$  = predicted data point using the distribution and its respective parameter(s)

$$\text{AIC} = 2k - 2\ln(\hat{L}) \quad (4)$$

where,

$k$  = number of estimated parameters

$\hat{L}$  = maximized value of the likelihood function for the distribution

$$\text{BIC} = k \ln(n) - 2\ln(\hat{L}) \quad (5)$$

where,

$k$  = number of estimated parameters

$n$  = number of observations (sample size)

$\hat{L}$  = maximized value of the likelihood function for the distribution

It is important to notice that the Fitter library follows the following rule to make its code more efficient: if a certain distribution takes more than 30 seconds to converge, it is assumed that it is not a good fit for the empirical data and thus the given distribution is skipped. For high performance computers, this is not an issue since the probability that those skipped distributions indeed have a bad fit is virtually 100%. Nonetheless, for ordinary computers this could create the risk of skipping a distribution that could potentially have a good fit. The computer used for this code was a moderate-high performance one, namely Lenovo ThinkPad T470s | i7-7600U | 24 GB RAM | 512 GB | SSD. Yet, to further mitigate this risk, this code was run 4 times, each time after restarting the computer in order to have full CPU and RAM capacity.

After the use of the Fitter library, the 5 best distributions based on the measures RSS, AIC and BIC were further studied using the following measures and tests: 1) Root-mean-square error/deviation (RMSD), 2) Error/residual standard deviation ( $S_{res}$ ), 3) Kolmogorov-Smirnov test (KS test), 4) Anderson-Darling test (AD test), 5) Cramér-von Mises test (CM test), and 6) Kuiper's test (K test). In order to integrate the Probability Density Function (PDF) of the further-studied distribution for the aforementioned goodness-of-fit tests, the Python library Scipy Stats was utilised.

RMSD, Equation (6), was used to determine the extent to which the use of each distribution is accurate to predict the likelihood of a certain value given its empirical likelihood. On the other hand,  $S_{res}$ , Equation (7), was used to determine the variation in the residuals of the predicted likelihoods. For both measures, the lower the value the better.

$$\text{RMSD} = \frac{\sqrt{\sum_{t=1}^T (\hat{y}_t - y_t)^2}}{T} \quad (6)$$

where,

$\hat{y}_t$  = predicted value

$y_t$  = empirical value

$T$  = number of observations or observed times

$$S_{res} = \sqrt{\frac{\text{RMSD} - \sqrt{\sum_{t=1}^T (\hat{y}_t - y_t)^2}}{T}} \quad (7)$$

where,

$\hat{y}_t$  = predicted value

$y_t$  = empirical value

$T$  = number of observations or observed times

The KS test, AD test, CM test and K test were used to determine the goodness-of-fit of each distribution as well as the probability that its deviations come

through randomness (through the use of the p-value). For each statistical test, the null hypothesis ( $H_0$ ) was that the certain distribution and the empirical data come from the same distribution. Consequently, the alternative hypothesis ( $H_1$ ) was that the certain distribution and the empirical data do not come from the same distribution. The KS test is presumably the most famous test among the tests used in this research and its equations are given as:

$$K_n \equiv \sqrt{n} \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \tag{8}$$

$$p = \exp(-K_n^2) \tag{9}$$

where,

$n$  = number of observations

sup = supremum

$\hat{F}_n(x)$  = empirical cumulative distribution function (CDF)

$F(x)$  = theoretical CDF

$p$  = estimated p-value

The AD test, on the other hand, gives more weight to the distribution tails than the KS test in such a way that the test is as sensitive in the tails as at the median [25]; and its equations are given as:

$$AD^2 = -N - S \tag{10}$$

$$S = \sum_{i=1}^N \frac{2i-1}{N} (\ln(x_i) + \ln(1-x_{N-i+1})) \tag{11}$$

$$p = e^{1.2937-5.709AD+0.0186AD^2} \text{ if } AD \geq 0.6 \tag{12}$$

where,

$N$  and  $n$  = number of observations

$x_i$  = predicted likelihood value from the theoretical CDF

The K test also gives more weight to the distribution tails as the AD test, yet it also makes itself invariant under cyclic transformations of the independent variable [26]. Its equations are given as:

$$D^+ = \frac{i}{n} - F(x_i) \tag{13}$$

$$D^- = F(x_i) - \frac{i-1}{n} \tag{14}$$

$$V = (\max(D^+) + \max(D^-))\sqrt{n} \tag{15}$$

$$p = \sum_{t=1}^{\infty} 2(4t^2V^2 - 1)e^{-2t^2V^2} \tag{16}$$

where,

$F(x_i)$  = theoretical CDF

$n$  = number of observations

Presumably the least well-known of all these tests, yet very powerful, the CM test equations are given as:

$$T = \frac{1}{12n} + \sum_{i=1}^n \left( \frac{2i-1}{2m} - F(x_i) \right)^2 \quad (17)$$

$$\sigma^2 \approx \frac{1}{45}$$

$$Z = \frac{T}{\sigma} \quad (18)$$

where,

$F(x_i)$  = theoretical CDF

$n$  = number of observations

$Z$  = z-score to be used in the two tail T-test in order to estimate the p-value

Lastly, to account for the fact that many trials were performed to find the best distribution, an adjusted Bonferroni correction was used. The classical Bonferroni correction is given by Equation (19) and states that there is a linear relationship between the number of trials and the significance level (this is proven by Equation (20)).

$$\alpha^* = \frac{\alpha}{\kappa} \quad (19)$$

$$\text{FWER} = P \left\{ \bigcup_{i=1}^{k_0} \left( p_i \leq \frac{\alpha}{k} \right) \right\} \leq \sum_{i=1}^{k_0} \left\{ P \left( p_i \leq \frac{\alpha}{k} \right) \right\} = k_0 \frac{\alpha}{k} \leq \alpha \quad (20)$$

where,

$\alpha$  = original significance level

$\alpha^*$  = new significance level

$k$  = number of trials

$p_i$  = p-value

Nevertheless, given that all tests  $H_0$  is that the certain distribution and the empirical data come from the same distribution, the use of the original Bonferroni correction would not be optimal. This is the case since the original Bonferroni decreases the Type-I error. However, for these goodness-of-fit tests, the Type-II error should be decreased because we are interested in the best fit for the empirical data and our main concern is to avoid the acceptance of the  $H_0$  when  $H_1$  is true. In other words, our main interest is to find the *best* distribution for the empirical data. As a result, the following formula was devised by the author:

$$\hat{p} = \frac{p}{k} \quad (21)$$

$\hat{p}$  = new estimated p-value of the certain distribution at a certain test

$p$  = estimated p-value of the certain distribution at a certain test

$k$  = number of trials (distributions) used

As a consequence, the p-values thresholds of every goodness-of-fit test are maintained and only the estimated distribution p-value is changed. That is, if our p-value threshold for KS test was 5%, it will continue to be 5% after this adjusted Bonferroni correction.

The mathematical proof thereof is given below:

Given Bonferroni's  $H_0$  evaluation equation:

$$p \leq \frac{\alpha}{k} \quad (22)$$

where,

$p$  = estimated p-value of the certain distribution at a certain test

$\alpha$  = confidence interval/p-value threshold

$k$  = number of trials (distributions) used

Let  $p_i$  and  $\alpha$  be constant. Given that  $k$  is used in Equation (22) to reduce the Type-I error and that there exists a negative linear relationship between Type-I error and Type-II error, then the only manner to decrease Type-II error by modifying Equation (22) is:

$$p \leq \alpha k \Leftrightarrow \frac{p}{k} \leq \alpha$$

The use of such a methodology with nine different goodness-of-fit measures plus the use of the adjusted Bonferroni correction provides a very thorough and solid method to assess the goodness-of-fit of various distribution in a certain random variable. This is the case since although all goodness-of-fit measures have their advantages, disadvantages and flaws, the combination of all of them compensate their disadvantages and flaws and thus gives a very complete picture of reality. For instance, the KS test is relatively more sensitive to values close to the mean, whereas the AD test is relatively more sensitive to the values in the tails [25], but the combination of both would give one a very good picture of all percentiles.

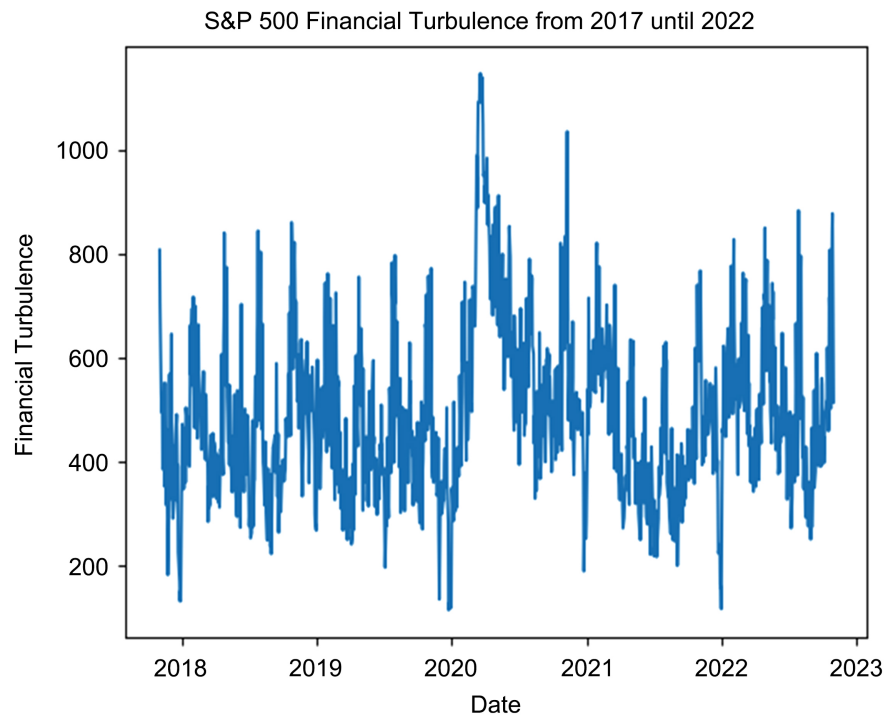
### 3. Results

The time-series development of the S&P 500 FT in the in-sample period can be found in **Figure 1**. As already expected, there is an exceptional rise in March and April 2020 (due to the first COVID lockdown in most countries in the world). In **Figure 2** one can find the histogram for the in-sample data. A right skewness and some excess kurtosis, when compared to the Gaussian distribution, can be easily spotted. This is confirmed in **Table 1**.

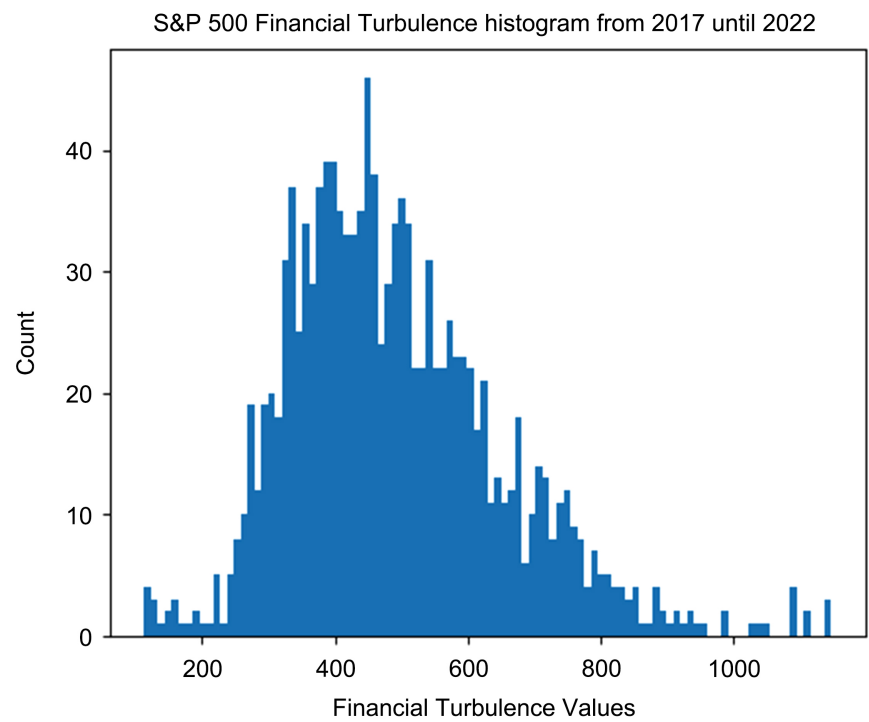
**Table 1.** S&P 500 FT in-sample basic statistics.

S&P 500 FT in-sample basic statistics	
Count	1257
Mean	492.00
Median	465.49
Standard Deviation	161.95
Minimum	112.95
Maximum	1146.71
Skewness	0.82
Excess Kurtosis	1.13



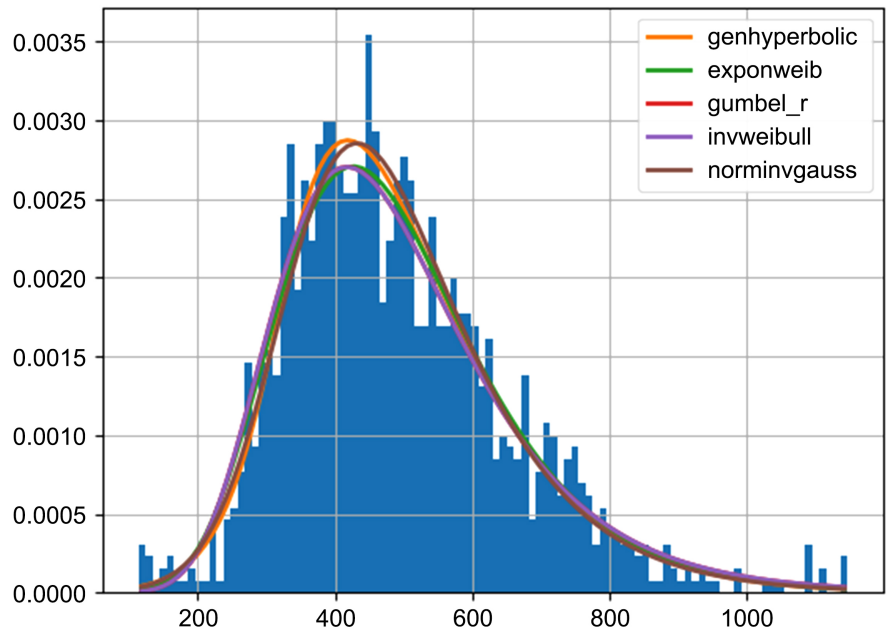


**Figure 1.** S&P 500 FT in-sample time-series development.

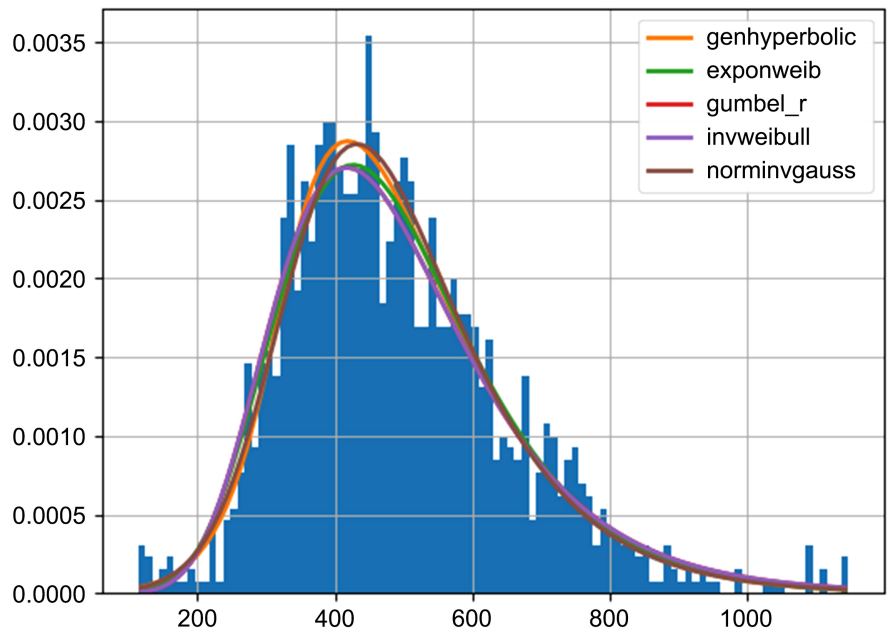


**Figure 2.** S&P 500 FT in-sample histogram.

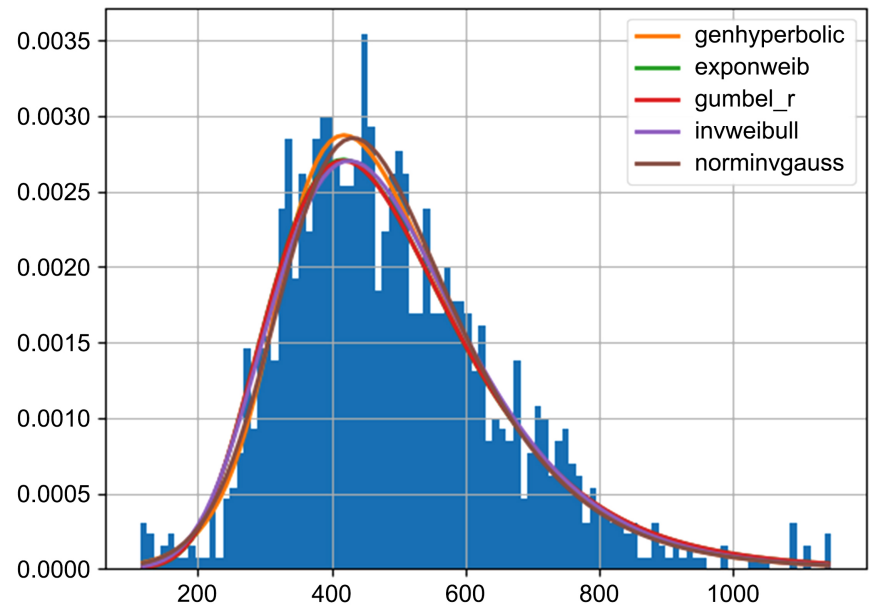
Given the results above, distributions that allow asymmetry should presumably be the ones to be studied in-depth. In **Figure 3**, the results of Python library Fitter can be found.



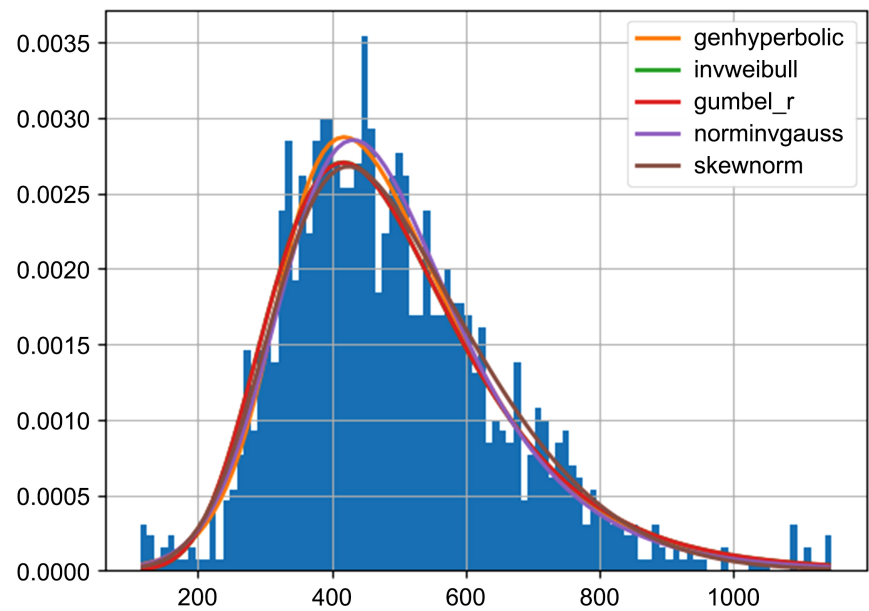
	sumsquare_error	aic	bic	kl_div	ks_statistic	ks_pvalue
genhyperbolic	0.000006	1559.210049	-24021.104692	inf	1.000000	0.000000
exponweib	0.000007	1555.506053	-23940.859629	inf	0.018355	0.783610
gumbel_r	0.000007	1545.271566	-23951.197205	inf	0.021683	0.587975
invweibull	0.000007	1547.130550	-23943.370681	inf	0.021709	0.586446
norminvgauss	0.000007	1557.403971	-23926.709907	inf	0.017003	0.854267



	sumsquare_error	aic	bic	kl_div	ks_statistic	ks_pvalue
genhyperbolic	0.000006	1559.205639	-24021.109189	inf	1.000000	0.000000
exponweib	0.000007	1558.082390	-23949.821255	inf	0.018023	0.801828
gumbel_r	0.000007	1545.271816	-23951.201973	inf	0.021683	0.587971
invweibull	0.000007	1547.254385	-23943.673129	inf	0.021741	0.584511
norminvgauss	0.000007	1557.404074	-23926.707971	inf	0.017000	0.854419



	sumsquare_error	aic	bic	kl_div	ks_statistic	ks_pvalue
genhyperbolic	0.000006	1559.208193	-24021.104359	inf	1.000000	0.000000
invweibull	0.000007	1548.353162	-23948.980436	inf	0.021514	0.597973
gumbel_r	0.000007	1545.271413	-23951.193331	inf	0.021682	0.588009
exponweib	0.000007	1555.252206	-23932.324014	inf	0.020096	0.682569
norminvgauss	0.000007	1557.403868	-23926.712262	inf	0.017000	0.854438



	sumsquare_error	aic	bic	kl_div	ks_statistic	ks_pvalue
genhyperbolic	0.000006	1559.210783	-24021.103129	inf	1.000000	0.000000
invweibull	0.000007	1547.545038	-23945.345708	inf	0.021637	0.590691
gumbel_r	0.000007	1545.271524	-23951.195688	inf	0.021681	0.588094
norminvgauss	0.000007	1557.403905	-23926.710981	inf	0.016997	0.854570
skewnorm	0.000007	1567.434956	-23931.976342	inf	0.020834	0.638475

Figure 3. Fitter library 4 tests results.

The distributions that appear in **Figure 3** are: 1) Generalised Hyperbolic distribution (GH), 2) Exponential Weibull distribution (EW), 3) Right-skewed Gumbel distribution (RG), 4) Inverse Weibull distribution (IW), 5) Normal-inverse Gaussian distribution (NIG), 6) Skew normal distribution (SN). Clearly, the results of each test converge; showing that the distributions presented in **Figure 3** are quite probably the best fit for the in-sample data. However, in the last test, there is a presence of a new distribution, namely the SN. This has probably occurred due to the 30-second safety rule of the algorithm that skipped the EW in this test. The striking result is that the GH was always the best fit in all tests. Despite this result, all distributions were further studied by passing through the aforementioned goodness-of-fit tests and error measures.

The respective parameters of each distribution for the in-sample data can be found in **Table 2**.

When performing their respective PDFs integrations with the Scipy Stats library, it was noticed that the integral of the GH PDF is probably divergent, or slowly convergent, given its respective parameters. Thus, as proposed by Hammerstein [27], the GH PDF could be approximated by Variance-Gamma distribution (VG) given its parameters. The mathematical proof thereof is given as:

Given  $\lambda > 0$  and the following norming constant (Equation (25)) obtained through the GH Lebesgue density (Equation (24)) given the GH's definition (Equation (23)):

**Table 2.** Distributions parameters.

GH	$\lambda = 2.7694$ $\alpha = 2.9027e-06$ $\beta = 1.3036e-06$ $\delta(\text{scale}) = 0.1466e-3$ $\mu(\text{location}) = 334.6479$
EW (as given [29])	$a = 1823.9813$ $c = 1.5987$ $\lambda(\text{scale}) = 465.4781$ $\mu(\text{location}) = -1222.1413$
RG	$\beta(\text{scale}) = 136.0440$ $\mu(\text{location}) = 417.0094$
IW (as given [30])	$c = 197203461.6252$ $\lambda(\text{scale}) = 26809877229.5768$ $\mu(\text{location}) = -26809876812.6089$
NIG	$\alpha = 4.7712$ $\beta = 2.9777$ $\delta(\text{scale}) = 246.4199$ $\mu(\text{location}) = 295.1719$
SN	$\alpha = 2.9635$ $\omega(\text{scale}) = 244.9482$ $\xi(\text{location}) = 308.1676$

$$GH(\lambda, \alpha, \beta, \delta, \mu) := N(\mu + \beta y, y) \circ GIG\left(\lambda, \delta, \sqrt{\alpha^2 + \beta^2}\right) \quad (23)$$

$$\begin{aligned} d_{GH(\lambda, \alpha, \beta, \delta, \mu)}(x) &= \int_0^{\infty} d_{N(\mu + \beta y, y)}(x) d_{GIG(\lambda, \delta, \sqrt{\alpha^2 + \beta^2})}(y) dy \\ &= a(\lambda, \alpha, \beta, \delta, \mu) \left( \delta^2 + (x - \mu)^2 \right)^{\frac{\lambda - 0.5}{2}} K_{\frac{\lambda - 1}{2}} \left( \alpha \sqrt{\delta^2 + (x - \mu)^2} \right) e^{\beta(x - \mu)} \end{aligned} \quad (24)$$

$$a(\lambda, \alpha, \beta, \delta, \mu) = \frac{(\alpha^2 - \beta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi} \alpha^{\lambda - \frac{1}{2}} \delta^{\lambda} K_{\lambda}(\delta \sqrt{\alpha^2 - \beta^2})} \quad (25)$$

Then:

$$a(\lambda, \alpha, \beta, \delta, \mu) \sim \frac{(\alpha^2 - \beta^2)^{\lambda}}{\sqrt{2\pi} \alpha^{\lambda - \frac{1}{2}} 2^{\lambda - 1} \Gamma(\lambda)} \text{ if } \delta \rightarrow 0 \text{ or } |\beta| \rightarrow \alpha$$

Given the GH  $\delta$ ,  $\alpha$ , and  $\beta$  values, this could be considered a good approximation. Also, given that  $\delta \rightarrow 0$ , then in Equation (23), we can conclude:

$$\sqrt{\delta^2 + (x - \mu)^2} \rightarrow |x - \mu|$$

Adding this to Equation (23), we get:

$$\begin{aligned} \lim_{\delta \rightarrow 0} d_{GH(\lambda, \alpha, \beta, \delta, \mu)}(x) &= \frac{(\alpha^2 - \beta^2)^{\lambda}}{\sqrt{2\pi} \alpha^{\lambda - \frac{1}{2}} 2^{\lambda - 1} \Gamma(\lambda)} |x - \mu|^{\lambda - 0.5} K_{\frac{\lambda - 1}{2}}(\alpha |x - \mu|) e^{\beta(x - \mu)} \\ &=: d_{VG(\lambda, \alpha, \beta, \mu)} \end{aligned}$$

Due to the fact that Scipy Stats library does not support the VG, the algorithm proposed by Laptev [28] was utilised. Yet, anew it was noticed that the integral of the VG PDF is probably divergent, or slowly convergent, given the parameters. Therefore, a numerical integration was performed by the author by using a loop for the GH PDF from 0 to 2000 with step = 1, followed by a calculation of the respective CDF value for each data point in the in-sample.

With the theoretical CDFs of each distribution and the empirical distribution of the in-sample, their respective RMSD and  $S_{res}$  were calculated and can be found in **Table 3**.

As already expected, the GH had the best results and the SN the worst ones. Nevertheless, the NIG yielded interesting results. Assuming  $\epsilon \sim N(\mu, \sigma^2)$  (where  $\epsilon$  = residual/error), then the NIG had even more precision than the EW.

In **Table 4** one can find all results of the aforementioned goodness-of-fit tests, with their respective p-values and adjusted Bonferroni corrected p-values ( $p_b$ ), where  $k = 6$  distributions. For the  $H_0$  and  $H_1$  evaluation, the following p-value thresholds (PVT) to *not* reject the  $H_0$  shall be used: 1) Conservative: p-value  $\geq 10\%$ , 2) Moderate: p-value  $\geq 5\%$ , 3) Relaxed: p-value  $\geq 1\%$ .

**Table 3.** Distributions RMSD and  $S_{res}$ .

GH	RMSD = 0.49337% $S_{res}$ = 0.2580%
EW	RMSD = 0.73304% $S_{res}$ = 0.4078%
RG	RMSD = 0.90224% $S_{res}$ = 0.5024%
IW	RMSD = 0.89799% $S_{res}$ = 0.5009%
NIG	RMSD = 0.741877% $S_{res}$ = 0.3616%
SN	RMSD = 0.96481% $S_{res}$ = 0.4902%

**Table 4.** Goodness-of-fit tests results.

	KS test	AD test	K test	CM test
GH	$K_n = 0.435$ $p = 82.75\%$ $p_b = 13.79\%$	$AD = 0.54$ $p = 17.08\%$ $p_b = 2.85\%$	$V = 0.857$ $p = 95.23\%$ $p_b = 15.87\%$	$T = 0.03$ $p = 83.46\%$ $p_b = 13.91\%$
EW	$K_n = 0.641$ $p = 66.29\%$ $p_b = 11.05\%$	$AD = 0.90$ $p = 2.22\%$ $p_b = 0.37\%$	$V = 1.201$ $p = 53.39\%$ $p_b = 8.90\%$	$T = 0.07$ $p = 65.48\%$ $p_b = 10.91\%$
RG	$K_n = 0.741$ $p = 57.78\%$ $p_b = 9.63\%$	$AD = 1.17$ $p = 0.48\%$ $p_b = 0.08\%$	$V = 1.386$ $p = 28.67\%$ $p_b = 4.78\%$	$T = 0.10$ $p = 48.35\%$ $p_b = 8.06\%$
IW	$K_n = 0.739$ $p = 57.92\%$ $p_b = 9.65\%$	$AD = 1.16$ $p = 0.49\%$ $p_b = 0.08\%$	$V = 1.379$ $p = 29.41\%$ $p_b = 4.90\%$	$T = 0.10$ $p = 48.70\%$ $p_b = 8.12\%$
NIG	$K_n = 0.603$ $p = 69.53\%$ $p_b = 11.59\%$	$AD = 0.74$ $p = 5.39\%$ $p_b = 0.90\%$	$V = 1.092$ $p = 69.76\%$ $p_b = 11.63\%$	$T = 0.07$ $p = 64.42\%$ $p_b = 10.74\%$
SN	$K_n = 0.739$ $p = 57.95\%$ $p_b = 9.66\%$	$AD = 0.93$ $p = 1.87\%$ $p_b = 0.31\%$	$V = 1.280$ $p = 41.95\%$ $p_b = 6.99\%$	$T = 0.11$ $p = 45.11\%$ $p_b = 7.52\%$

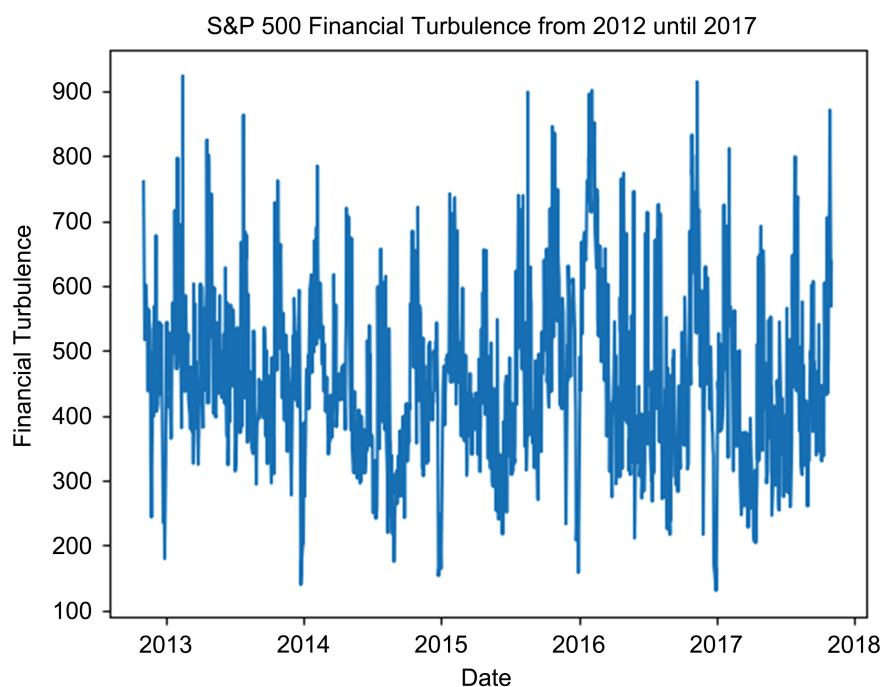
As expected, the GH had the best results once more, passing all tests considering the Conservative PVT besides the AD test with the  $p_b$  where it only passed this test considering the Relaxed PVT. This can be interpreted as the following: the GH represents the in-sample data with high accuracy, yet when looking at its tails, the GH fails to adequately represent it. Again, the NIC had

better results than all the other distributions, something striking given its scores in the Fitter library tests. Another interesting result is that the SN outperformed the RG and IW in the AD and K tests and had similar results in the KS test. In conclusion, for the in-sample data, the GH was by far the best fit. However, to test these results, this whole process will be repeated for the out-sample data.

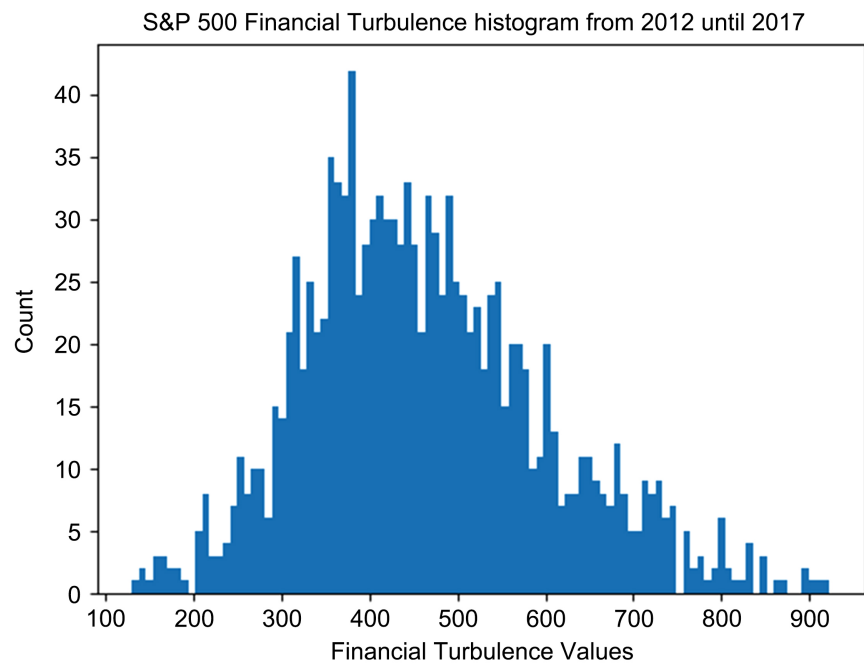
The time-series development of the S&P 500 FT in the out-sample period can be found in **Figure 4**. As already expected, there is no exceptional rise as the one in March and April 2020. In **Figure 5** one can find the histogram for the out-sample data. Again, a right skewness and some excess kurtosis can be spotted, albeit less than the in-sample data. This is confirmed in **Table 5**.

**Table 5.** S&P 500 FT out-sample basic statistics.

S&P 500 FT out-sample basic statistics	
Count	1258
Mean	464.00
Median	447.31
Standard Deviation	137.22
Minimum	130.26
Maximum	923.21
Skewness	0.51
Excess Kurtosis	0.14



**Figure 4.** S&P 500 FT out-sample time-series development.



**Figure 5.** S&P 500 FT out-sample histogram.

In **Figure 6**, the results of Python library Fitter for the studied distributions can be found:

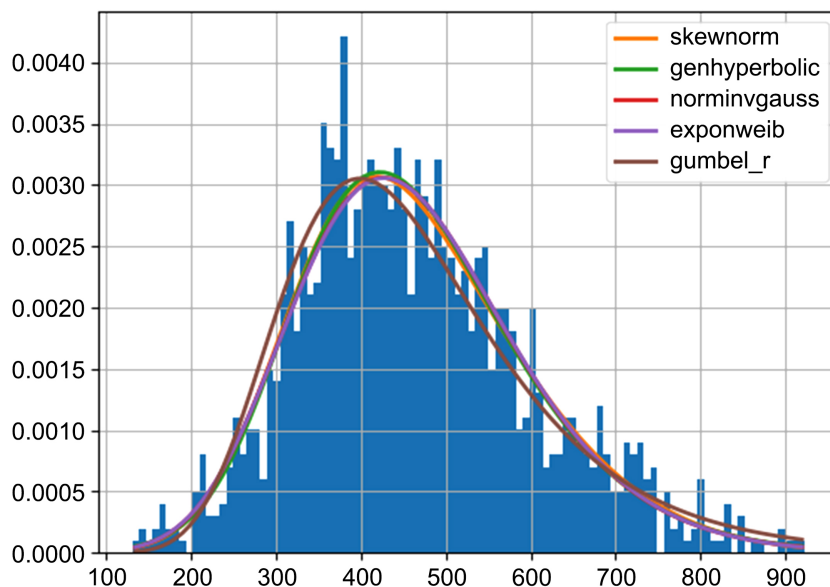
The results show a less accurate fit of the distributions than the in-sample. However, the GH is still high classified. The respective parameters of each distribution for the out-sample data can be found in **Table 6**. This time, it was possible to integrate all PDFs through the use of the Scipy Stats library given their respective parameters. With the theoretical CDFs of each distribution and the empirical distribution of the out-sample, their respective RMSD and  $S_{res}$  were calculated and can be found in **Table 7**.

As already predicted, all distributions besides the SN had more inaccuracy than the in-sample. However, the GH still has the best results when disregarding the SN.

In **Table 8** on the other hand, one can find all results of the aforementioned goodness-of-fit tests, with their respective p-values and adjusted Bonferroni corrected p-values ( $p_b$ ), where  $k = 6$  distributions. For the  $H_0$  and  $H_1$  evaluation, the same PVT to *not* reject the  $H_0$  was used as in the in-sample part.

Anew, all distributions besides the SN had worse results. Yet, GH had the best results again when not considering the SN. It passed all tests considering the Conservative PVT besides the AD tests. Considering the “normal” p-value, it did not pass the AD test only when considering the Conservative PVT. On the other hand, regarding the  $p_b$  it only passed the AD test considering the Relaxed PVT. This can be interpreted as the following: the GH also represents the out-sample data accurately, yet when looking at its tails, the GH fails to represent it properly. Again, the NIC had better results than EW, RG and IW. Finally, in the out-sample data, the SN overperformed all distributions.





	sumsquare_error	aic	bic	kl_div	ks_statistic	ks_pvalue
skewnorm	0.000011	1460.097497	-23357.707799	inf	0.015245	0.927452
genhyperbolic	0.000011	1464.125426	-23337.251950	inf	0.018047	0.800150
norminvgauss	0.000011	1462.312698	-23290.212139	inf	0.018871	0.754021
exponweib	0.000011	1462.134574	-23284.796167	inf	0.019545	0.714788
gumbel_r	0.000012	1456.662635	-23244.951472	inf	0.029556	0.217463

Figure 6. Fitter library results.

Table 6. Distribution parameters.

GH	$\lambda = 5.814779919765607$ $\alpha = 0.0766758193638856$ $\beta = 0.03225410918556625$ $\delta(\text{scale}) = 2.350218464596915$ $\mu(\text{location}) = 281.80398210965654$
EW (as parametrized by [29])	$a = 36.7987564428992$ $c = 2.510118936749314$ $\lambda(\text{scale}) = 671.2349848193721$ $\mu(\text{location}) = -712.507095588652$
RG	$\beta(\text{scale}) = 120.41477425095212$ $\mu(\text{location}) = 398.7084109936449$
IW (as parametrized by [30])	$c = 208417228.78456452$ $\lambda(\text{scale}) = 25247906954.219852$ $\mu(\text{location}) = -25247906555.182213$
NIG	$\alpha = 20.070176743758395$ $\beta = 14.574033088198112$ $\delta(\text{scale}) = 351.29588934780935$ $\mu(\text{location}) = 92.97049346881448$
SN	$\alpha = 2.377525133892676$ $\omega(\text{scale}) = 202.40715218591686$ $\xi(\text{location}) = 315.1567034804059$

**Table 7.** Distributions RMSD and  $S_{res}$ .

GH	RMSD = 0.68313% $S_{res}$ = 0.4209%
EW	RMSD = 0.85291% $S_{res}$ = 0.4687%
RG	RMSD = 1.42144% $S_{res}$ = 0.6330%
IW	RMSD = 1.41006% $S_{res}$ = 0.6422%
NIG	RMSD = 0.820871% $S_{res}$ = 0.4543%
SN	RMSD = 0.63393% $S_{res}$ = 0.3445%

**Table 8.** Goodness-of-fit results.

	KS test	AD test	K test	CM test
GH	$K_n = 0.640$ $p = 66.38\%$ $p_b = 11.06\%$	$AD = 0.67$ $p = 7.96\%$ $p_b = 1.33\%$	$V = 1.153$ $p = 60.64\%$ $p_b = 10.11\%$	$T = 0.06$ $p = 69.79\%$ $p_b = 11.63\%$
EW	$K_n = 0.693$ $p = 61.85\%$ $p_b = 10.31\%$	$AD = 0.80$ $p = 3.76\%$ $p_b = 0.63\%$	$V = 1.237$ $p = 48.07\%$ $p_b = 8.01\%$	$T = 0.09$ $p = 54.71\%$ $p_b = 9.12\%$
RG	$K_n = 1.020$ $p = 35.31\%$ $p_b = 5.89\%$	$AD = 1.57$ $p = 0.05\%$ $p_b = 0.01\%$	$V = 1.898$ $p = 1.99\%$ $p_b = 0.33\%$	$T = 0.26$ $p = 8.10\%$ $p_b = 1.35\%$
IW	$K_n = 1.016$ $p = 35.65\%$ $p_b = 5.94\%$	$AD = 1.58$ $p = 0.05\%$ $p_b = 0.01\%$	$V = 1.931$ $p = 1.61\%$ $p_b = 0.27\%$	$T = 0.25$ $p = 8.73\%$ $p_b = 1.45\%$
NIG	$K_n = 0.669$ $p = 63.90\%$ $p_b = 10.65\%$	$AD = 0.79$ $p = 4.13\%$ $p_b = 0.69\%$	$V = 1.207$ $p = 52.49\%$ $p_b = 8.75\%$	$T = 0.08$ $p = 57.62\%$ $p_b = 9.60\%$
SN	$K_n = 0.541$ $p = 74.65\%$ $p_b = 12.44\%$	$AD = 0.64$ $p = 9.76\%$ $p_b = 1.63\%$	$V = 1.021$ $p = 79.59\%$ $p_b = 13.27\%$	$T = 0.05$ $p = 73.92\%$ $p_b = 12.32\%$

In conclusion, it can be affirmed with quite some confidence that at least in the past 10 years, the S&P 500 FT followed a GH, though there may exist some deviations from this distribution in high and low quantiles. This result is not remarkable given the fairly common presence of the GH in the financial research field [31] [32] [33] [34].

## 4. Conclusions

The objective of this study was to determine the best distribution for the S&P 500 FT such that researchers can further study this financial phenomenon as well as allow asset or risk managers from financial institutions to make a more accurate use of this financial phenomenon in their risk and portfolio performance analyses. This objective was achieved through the use of the various statistical measures and tests. These being: 1) RSS, 2) AIC, 3) BIC, 4) RMSD, 5)  $S_{res}$ , 6) KS test, 7) AD test, 8) CM test, and 9) K test. Additionally, the use of in-sample and out-sample data as well as an adjusted Bonferroni correction was made in order to increase the reliability of the results.

As a result of these methods, it was found that the GH was the best distribution to describe this financial phenomenon for the S&P 500. It has remarkably lower prediction errors than the other studied distributions (EW, RG, IW, NIG and SN) when both the in-sample and out-sample results are considered. Furthermore, for both the in-sample and out-sample, the GH passed all goodness-of-fit tests considering the Conservative PVT, besides the most AD tests where it only passed considering the Moderate and/or Relaxed PVT. Thus, it can be confidently concluded that the GH was the best distribution for the S&P 500 FT in the last 10 years. Furthermore, if the future American financial market does not change to a great extent, the GH will theoretically continue to be the best distribution for the S&P 500 FT.

In conclusion, researchers and financial institutions can assume that the S&P 500 FT follows the GH with quite some precision when making use of this risk parameter in their studies, simulations or analyses. Nevertheless, they should not forget that the S&P 500 FT behaviour could theoretically change in the future due to major changes in the American and global economy or financial market. Consequently, the GH might not be the best fit distribution for the S&P 500 FT anymore. Therefore, ideally researchers and financial institutions should periodically (e.g. every 5 years) perform a best fit distribution study as the one performed in this paper in order to have the most accurate results in their studies, simulations or analyses. On top of that, as Zhao and Li [35] (unfortunately after the devise of this paper's methodology) showed that the use of modified goodness-of-fit tests with the use of best linear unbiased scale (BLUS) could potentially provide more accurate results. Therefore, one could also use BLUS in their future best fit distribution studies.

Lastly, the author would like to again encourage the scientific community to perform a similar study as this one with the use of BLUS for other indexes and markets FT, and as well as for the other aforementioned risk parameter, Absorption Ratio.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- [1] Ang, A. and Chen, J. (2002) Asymmetric Correlations of Equity Portfolios. *Journal of Financial Economics*, **63**, 443-494. [https://doi.org/10.1016/S0304-405X\(02\)00068-5](https://doi.org/10.1016/S0304-405X(02)00068-5)
- [2] Malkiel, B.G. (1973) *A Random Walk down Wall Street: The Time-Tested Strategy for Successful Investing*. 11th Edition, W. W. Norton & Company, New York.
- [3] Demirer, R., Gupta, R., Lv, Z. and Wong, W.K. (2019) Equity Return Dispersion and Stock Market Volatility: Evidence from Multivariate Linear and Nonlinear Causality Tests. *Sustainability*, **11**, 351. <https://doi.org/10.3390/su11020351>
- [4] Engle, R.F., Ghysels, E. and Sohn, B. (2013) Stock Market Volatility and Macroeconomic Fundamentals. *Review of Economics and Statistics*, **95**, 776-797. [https://doi.org/10.1162/REST\\_a\\_00300](https://doi.org/10.1162/REST_a_00300)
- [5] Inci, A.C., Li, H. and McCarthy, J. (2011) Financial Contagion: A Local Correlation Analysis. *Research in International Business and Finance*, **25**, 11-25. <https://doi.org/10.1016/j.ribaf.2010.05.002>
- [6] Liu, R., Demirer, R., Gupta, R. and Wohar, M. (2019) Volatility Forecasting with Bivariate Multifractal Models. *Journal of Forecasting*, **39**, 155-167. <https://doi.org/10.1002/for.2619>
- [7] Poon, S. and Clive, W.J. (2003) Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature*, **41**, 478-539. <https://doi.org/10.1257/jel.41.2.478>
- [8] Rangel, J.G. and Engle, R.F. (2012) The Factor-Spline-GARCH Model for High and Low Frequency Correlations. *Journal of Business & Economic Statistics*, **30**, 109-124. <https://doi.org/10.1080/07350015.2012.643132>
- [9] Salisu, A.A. and Gupta, R. (2021) Oil Shocks and Stock Market Volatility of the BRICS: A GARCH-MIDAS Approach. *Global Finance Journal*, **48**, Article ID: 100546. <https://doi.org/10.1016/j.gfj.2020.100546>
- [10] Salisu, A.A. and Ogbonna, A.E. (2022) The Return Volatility of Cryptocurrencies during the COVID-19 Pandemic: Assessing the News Effect. *Global Finance Journal*, **54**, Article ID: 100641. <https://doi.org/10.1016/j.gfj.2021.100641>
- [11] Kritzman, M. and Li, Y. (2010) Skulls, Financial Turbulence, and Risk Management. *Financial Analysts Journal*, **66**, 30-41. <https://doi.org/10.2469/faj.v66.n5.3>
- [12] Kritzman, M., Li, Y., Page, S. and Rigobon, R. (2011) Principal Components as a Measure of Systemic Risk. *The Journal of Portfolio Management*, **37**, 112-126. <https://doi.org/10.3905/jpm.2011.37.4.112>
- [13] Salisu, A.A., Demirer, R. and Gupta, R. (2022) Financial Turbulence, Systemic Risk and the Predictability of Stock Market Volatility. *Global Finance Journal*, **52**, Article ID: 100699. <https://doi.org/10.1016/j.gfj.2022.100699>
- [14] Nystrup, P., Boyd, S., Lindström, E. and Madsen, H. (2018) Multi-Period Portfolio Selection with Drawdown Control. *Annals of Operations Research*, **282**, 245-271. <https://doi.org/10.1007/s10479-018-2947-3>
- [15] Liu, X.Y., Yang, H., Gao, J. and Wang, C. (2021) FinRL: Deep Reinforcement Learning Framework to Automate Trading in Quantitative Finance. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3955949>
- [16] Chan, S., Chu, J., Nadarajah, S. and Osterrieder, J. (2017) A Statistical Analysis of Cryptocurrencies. *Journal of Risk and Financial Management*, **10**, 12. <https://doi.org/10.3390/jrfm10020012>

- [17] Egan, W.J. (2007) The Distribution of S&P 500 Index Returns. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.955639>
- [18] Ignatieva, K. and Landsman, Z. (2019) Conditional Tail Risk Measures for the Skewed Generalised Hyperbolic Family. *Insurance: Mathematics and Economics*, **86**, 98-114. <https://doi.org/10.1016/j.insmatheco.2019.02.008>
- [19] McNeil, A.J. (2002) The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance. *Journal of the American Statistical Association*, **97**, 1210-1211. <https://doi.org/10.1198/jasa.2002.s242>
- [20] Rydberg, T.H. (2000) Realistic Statistical Modelling of Financial Data. *International Statistical Review*, **68**, 233-258. <https://doi.org/10.1111/j.1751-5823.2000.tb00329.x>
- [21] Seneta, E. (2004) Fitting the Variance-Gamma Model to Financial Data. *Journal of Applied Probability*, **41**, 177-187. <https://doi.org/10.1239/jap/1082552198>
- [22] Brockett, P.L. (1983) On the Misuse of the Central Limit Theorem in Some Risk Calculations. *The Journal of Risk and Insurance*, **50**, 727. <https://doi.org/10.2307/252712>
- [23] Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013) Model Uncertainty and VaR Aggregation. *Journal of Banking & Finance*, **37**, 2750-2764. <https://doi.org/10.1016/j.jbankfin.2013.03.014>
- [24] Kinnebrock, S. and Podolskij, M. (2008) A Note on the Central Limit Theorem for Bipower Variation of General Functions. *Stochastic Processes and Their Applications*, **118**, 1056-1070. <https://doi.org/10.1016/j.spa.2007.07.009>
- [25] Saculinggan, M. and Balase, E.A. (2013) Empirical Power Comparison of Goodness of Fit Tests for Normality in the Presence of Outliers. *Journal of Physics: Conference Series*, **435**, Article ID: 012041. <https://doi.org/10.1088/1742-6596/435/1/012041>
- [26] Kuiper, N.H. (1960) Tests Concerning Random Points on a Circle. *Indagationes Mathematicae (Proceedings)*, **63**, 38-47. [https://doi.org/10.1016/S1385-7258\(60\)50006-0](https://doi.org/10.1016/S1385-7258(60)50006-0)
- [27] Frhr. v. Hammerstein, E.A. (2010) Generalized Hyperbolic Distributions: Theory and Applications to CDO Pricing. PhD Dissertation, Albert-Ludwigs-Universität, Freiburg.
- [28] Laptev, D. (2018, March 18) GitHub-dlaptev/VarGamma: Variance Gamma Distribution (Python): Pdf, cdf, rand and fit. GitHub. <https://github.com/dlaptev/VarGamma>
- [29] Scipy (n.d.) Scipy.stats.exponweib—SciPy v1.9.3 Manual. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.exponweib.html>
- [30] Scipy (n.d.) Scipy.stats.invweibull—SciPy v1.9.3 Manual. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.invweibull.html>
- [31] Eberlein, E. and Prause, K. (2002) The Generalized Hyperbolic Model: Financial Derivatives and Risk Measures. *The First World Congress of the Bachelier Finance Society*, Paris, 29 June-1 July 2000, 245-267. [https://doi.org/10.1007/978-3-662-12429-1\\_12](https://doi.org/10.1007/978-3-662-12429-1_12)
- [32] Fajardo, J. and Farias, A. (2004) Generalized Hyperbolic Distributions and Brazilian Data. *Brazilian Review of Econometrics*, **24**, 249. <https://doi.org/10.12660/bre.v24n22004.2712>
- [33] Prause, K. (1997) Modelling Financial Data Using Generalized Hyperbolic Distributions. Universität Freiburg, Freiburg.

- [34] Wang, C. (2005) On the Numerics of Estimating Generalized Hyperbolic Distribution. Master Thesis, Humboldt-Universität zu Berlin, Berlin.
- [35] Zhao, J. and Li, X. (2022) Goodness of Fit Test Based on BLUS Residuals for Error Distribution of Regression Model. *Applied Mathematics*, **13**, 672-682.  
<https://doi.org/10.4236/am.2022.138042>

## Appendix 1

["\_fit", "alpha", "anglit", "arcsine", "argus", "beta", "betaprime", "bradford", "burr", "burr12", "cauchy", "chi", "chi2", "cosine", "crystalball", "dgamma", "dweibull", "erlang", "expon", "exponnorm", "exponpow", "exponweib", "f", "fatiguelife", "fisk", "foldcauchy", "foldnorm", "gamma", "gausshyper", "genexpon", "genextreme", "gengamma", "genhalflogistic", "genhyperbolic", "geninvgauss", "genlogistic", "gennorm", "genpareto", "gibrat", "gilbrat", "gompertz", "gumbel\_l", "gumbel\_r", "halfcauchy", "halfgennorm", "halflogistic", "halfnorm", "hypsecant", "invgamma", "invgauss", "invweibull", "johnsonsb", "johnsonsu", "kappa3", "kappa4", "ksone", "kstwo", "kstwobign", "laplace", "laplace\_asymmetric", "levy", "levy\_l", "levy\_stable", "loggamma", "logistic", "loglaplace", "lognorm", "loguniform", "lomax", "maxwell", "mielke", "moyal", "nakagami", "ncf", "nct", "ncx2", "norm", "norminvgauss", "pareto", "pearson3", "powerlaw", "powerlognorm", "powernorm", "rayleigh", "rdist", "recipinvgauss", "reciprocal", "rice", "rv\_continuous", "rv\_histogram", "semicircular", "skewcauchy", "skewnorm", "studentized\_range", "t", "trapezoid", "trapz", "triang", "truncexpon", "truncnorm", "truncweibull\_min", "tukeylambda", "uniform", "vonmises", "vonmises\_line", "wald", "weibull\_max", "weibull\_min", "wrapcauchy"]