Historical Simulation systematically underestimates the Expected Shortfall

December 2, 2024

Pablo García-Risueño

University of Zaragoza, Spain.

risueno@unizar.es, garcia.risueno@gmail.com

Historical Simulation systematically underestimates the Expected Shortfall

December 2, 2024

Abstract

Expected Shortfall (ES) is acquiring an increasingly relevant role in financial risk management. ES is often calculated using Historical Simulation (HS); this has advantages like being parameter-free and has been favoured by some regulators. However, the usage of HS for calculating ES presents a potentially serious drawback: It strongly depends on the size of the sample of historical data. Moreover, this relationship leads to systematic underestimation: the lower the sample size, the lower the ES tends to be. In this letter we make a brief presentation of this phenomenon, and present some remarks on its impact in the analysis of both illiquid and liquid financial products.

Highlights:

- Historical Simulation is gaining relevance for risk measurement, partly because of regulatory requests.
- Calculation of the Expected Shortfall using Historical Simulation is unreliable unless the sample size is high enough, which can be problematic due to data staleness.
- It is possible to increase the accuracy of the calculated Expected Shortfall by using continuous functions fitted to historical data.

Keywords: Expected Shortfall, Historical Simulation, fat-tailed probability density functions, bonds, stocks, illiquidity.

JEL codes: C22, C54, C61, C87.

I. Introduction

Value-at-Risk (VaR, Choudhry (2013)) and Expected Shortfall (McNeil et al. (2005), ES) have gradually become quantities of the utmost importance in the area of risk management. VaR does not consider the severity of the potential losses beyond a given threshold. Moreover, it presents other drawbacks, like not being subadditive, which makes it not reflect the reduction of risk through diversification well (McNeil et al., 2005). Therefore the Expected Shortfall is acquiring an increasing relevance. ES (also called Conditional Value at Risk –CVaR–) presents other inconveniences, like being strictly speaking non-elicitable (Carrillo Menéndez and Hassani, 2021). Nevertheless, the fact that it gives more weight to more severe losses makes it a more descriptive and reliable risk measure. In this short article, we discuss methods to calculate ES accurately. We compare them to the method of Historical Simulation (HS), understood as *using a discrete set of observed returns*, without fitting them to any probability density function (pdf). Our analysis focuses on the amount of data employed to evaluate the ES. This is not the first time that this question has been investigated. Several authors in the past have researched the impact of the number of points of the sample (sample size) on different properties of the ES. For example, Liu and Staum (2010) evaluated the relative root mean squared error of the Expected Shortfall using distinct methods as a function of the number of simulated payoffs in the context of nested Monte Carlo simulation of portfolios; Wong (2008) analyzed the outliers of a normality test based on the ES as a function of the sample size; Maio et al. (2017) also calculated the ES using HS and other methods, yet only for two different sample sizes (1k and 100k points). Our analysis differs from theirs, being our goal to provide the community of financial Mathematics and risk management with insights on the features of the Expected Shortfall by providing a clear and intuitive presentation of them. We also suggest methods for the accurate calculation of ES. For the sake of simplicity and intuitiveness, we analyse single financial products rather than portfolios.

This article is structured as follows. In sec. *Data* we mention the financial products that we use for our analysis. The bulk of our research is presented in sec. *Methods and results*, which we present together for the sake of clarity. This section is divided into three subsections. In the first one, we describe how discrete datasets are fitted to probability density functions. In the second one, we display *How the Expected Shortfall from Historical Simulation depends on the size of the dataset*. In the third one, we propose a method for accurate ES calculations. Finally, in sec. *Conclusions* we make a summary of the whole research. More extensive results are presented in the Supporting Information. The code employed in our calculations is publicly available in the `github` repository of the first author (Garcia-Risueño, 2023).

II. Data

We have analyzed the time series of two bonds and two stocks, whose principal details are presented in Tab. 1. We have chosen bonds and stocks because they are among the most traded financial products. Moreover, they correspond to diverse sectors (chemistry, technology, finance and energy) and to different developed regions (North America and Europe). The time series of the bonds were downloaded from finanzen.net; the time series of the stocks were downloaded from Yahoo Finance (finance.yahoo.com). All the employed time series have daily periodicity. Our theses are a consequence of generic mathematical properties, not of specific features of financial products. Therefore we deem it unnecessary to tackle a higher number of time series.

ISIN	Kind of product	Company	Sector	Region (currency)	Time range
XS1017833242	bond	BASF, SE	Chemistry	Europe (€)	2014.03.20 ∼ 2023.06.10
US808513AL92	bond	Charles Schwab	Finance	North America ($)	2017.05.29 ∼ 2023.06.26
GB00BP6MXD84	stocks	Shell p.l.c.	Energy	Europe (€)	2018.06.22 ∼ 2023.06.22
US0378331005	stocks	Apple Inc.	Technology	North America ($)	2018.06.22 ∼ 2023.06.22

Table 1: List of financial products whose time series were used in this research.

We have calculated absolute returns for bonds and logarithmic returns for stocks. In the latter case, we considered reinvestment of dividends. Further explanations on the procedure to calculate the returns for stocks are presented in Ref. Garcia-Risueño et al. (2023). We did not consider reinvestment of the coupon payments of the analyzed bonds because, due to the fact that the accrued interest must be paid to the bond seller, the bond price does not abruptly change due to coupon payments. We define the logarithmic return of the price p^A of a stock A as the difference of the logarithms of the prices (Hafez and Lautizi, 2019), this is:

$$r^A(t_i) := r_i := \log\left[\frac{p^A(t_i)}{p^A(t_{i-1})}\right] = \log\left[p^A(t_i)\right] - \log\left[p^A(t_{i-1})\right] \quad . \tag{1}$$

where t and log represent the time and natural logarithm (ln), respectively; $(t-1)$ indicates the time of the time series immediately previous to t (in our analysis, t and $(t-1)$ differ in one trading day).

III. Methods and results

In this section, we present the manner we performed our calculations and their outcomes, in an attempt to reveal the effect of the sample size on the ES and how eventual inaccuracies might be mitigated.

III.A. Fitting to probability distributions

We have considered four different probability distributions, including three fat-tailed ones, to fit the time series of returns. These are the well-known normal distribution (Gaussian probability density function), the non-centered (and potentially skewed) t-student distribution, the generalized hyperbolic distribution, and the Lévy stable distribution. This set of distributions was chosen just for illustrating purposes, i.e. to present the properties and usefulness of the method which consists of fitting the returns to fat-tailed distributions. The finance practitioner can also include other distributions and select them for the calculation of the ES if the observed data fits them better.

For the normal distribution, the location (loc) and scale parameters are simply the average and standard deviation of the returns. Concerning the fat-tailed distributions, we fitted them using the maximum likelihood method, minimizing a loss function is defined as follows:

$$\text{loss}(\mathbf{p}_d) \; := \; -\frac{1}{N_r} \sum_{i=1}^{N_r} \ln[\, \text{pdf}(r_i; \mathbf{p}_d) \,] \tag{2}$$

where N_r stands for the number of datapoints (returns) and pdf is the tested probability density function. Therefore, the lower the loss function, the more accurate the fitting of the probability density function to the dataset. We call $\mathbf{p}_d$ the set of two, four, or five parameters of the distribution. The minimization of the loss function was carried out using the gradient descent method with step size à la Barzilai-Borwein (Barzilai and Borwein, 1988), including some randomness in the calculations (to reach different local minima of the loss function) and using different initial values of the parameters of the distributions. Further explanations of the fitting to distributions can be viewed in Garcia-Risueño et al. (2023); Johnson et al. (1995) or in the shared source code used in our calculations.

In Fig. 1 we present an example of the fitting of the analyzed probability density functions to one of the chosen datasets (returns of the BASF bond, in this case; similar plots for the other analyzed products are presented in the Supporting Information). In Fig. 1 the dark blue bars represent the histogram of the absolute returns of the bond price, and the continuous lines represent the probability density functions. In this case, the loss function is minimal for the generalized hyperbolic function. This happens in 3 out of the 4 analyzed products. The non-centered t-student has minimal loss function in the remaining case. The parameters from our fitting, as well as their corresponding loss functions, are presented in Tab. 2. In the Supporting Information we present example Q-Q plots for the analyzed probability distributions for all the analyzed financial products.

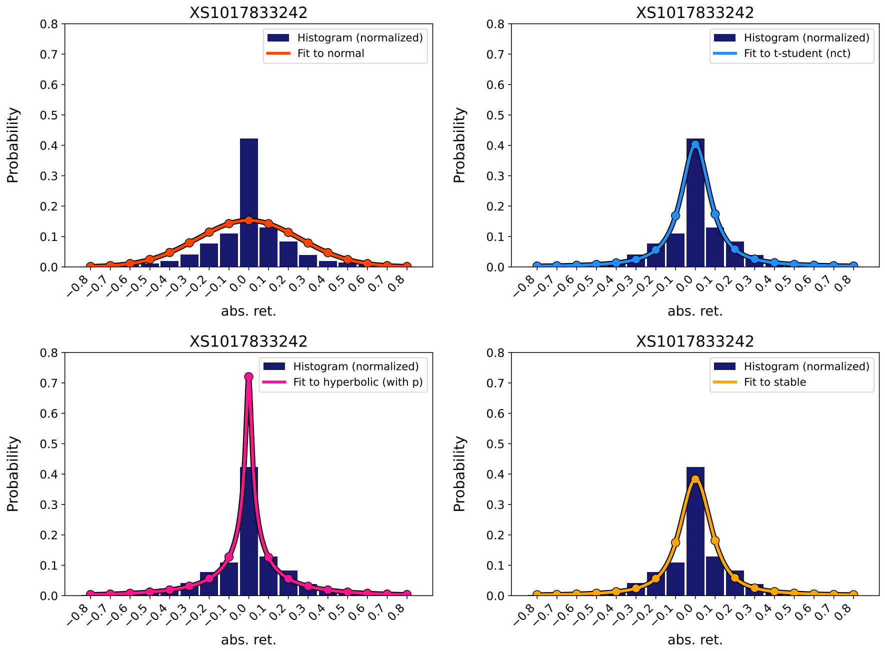

Figure 1: Fitting of observed absolute returns of the price of a bond (ISIN XS1017833242) –represented in the histogram– to different probability density functions. Top, left: Normal; Top, right: Non-centered t-student; Bottom, left: Generalized hyperbolic; Bottom, right: Lévy stable.

Product name	normal loc	normal scale	normal loss
BASF bond	-8.07 $\cdot 10^{-4}$	0.26052	0.073885
Charles Schwab bond	-2.68 $\cdot 10^{-3}$	0.664115	1.00963
Apple stocks	1.15 $\cdot 10^{-3}$	0.020996	-2.4445
Shell stocks	1.61 $\cdot 10^{-4}$	0.022109	-2.392818

Product name	nct loc	nct scale	nct sk. param	nct df	nct loss
BASF bond	-9.87 $\cdot 10^{-5}$	0.083185	0.017518	1.3453	-0.24973
Charles Schwab bond	-9.17 $\cdot 10^{-3}$	0.153745	0.02477200	1.56684	0.244798
Apple stocks	4.85 $\cdot 10^{-3}$	0.0146267	-0.20063	3.7456	**-2.522572**
Shell stocks	2.18 $\cdot 10^{-3}$	0.012391	-0.108819	2.66297	-2.569654

Product name	g. hyp. loc	g. hyp. scale	g. hyp. b param	g. hyp. a param	g. hyp. p param	g. hyp. loss
BASF bond	1.21 $\cdot 10^{-3}$	0.096572	-0.0035894	0.30363	-0.25482	**-0.25081**
Charles Schwab bond	-4.03 $\cdot 10^{-3}$	0.187008	0.000523468	0.0170110	-0.755003	**0.244552**
Apple stocks	2.11 $\cdot 10^{-3}$	0.024513	-0.054002	0.52974	-1.3768	-2.522517
Shell stocks	7.90 $\cdot 10^{-4}$	0.017955	-0.023476	0.240805	-1.0703240247426045	**-2.570094**

Product name	stable loc	stable scale	stable β	stable α	stable loss
BASF bond	1.55 $\cdot 10^{-5}$	0.074546	0.0069359	1.3196	-0.22360
Charles Schwab bond	9.06 $\cdot 10^{-3}$	0.13826	-0.012709	1.2849	0.2481565
Apple stocks	1.41 $\cdot 10^{-3}$	0.011747	-0.070499	1.6800	-2.5175
Shell stocks	2.45 $\cdot 10^{-4}$	0.010717	-0.110393	1.67947	-2.559161

Table 2: Parameters of the fitting of the returns to four probability distributions: normal, non-centered t-student (nct), generalized hyperbolic (g. hyp.) and stable. Bold font indicates the best fitting among all four analyzed distributions.

We will use the fitted distributions to generate synthetic data, which will be used for inferring properties of the ES of the returns of financial products. The usage of synthetic data has been encouraged by prestigious authors (López de Prado, 2018). In order to avoid too extreme values, we establish truncation values for the synthetic data. If the generated random numbers –which correspond to absolute returns of bonds– are above +30 or below -30 (note that the prices of bonds are usually measured so that they are about 100 currency units) then the generated synthetic datum is discarded, and another random number is generated. We deem these round values reasonable for bonds; for example, the bond with ISIN US33616CAB63 (First Republic Bank) fell over 23 in a single day in spring 2023. For the stocks of Shell we set maximum and minimum truncation limits of log(1.4) and log(0.6), which are consistent with historical extreme values of oil companies (+36% of BP on 1969-06-04 and -47% on 202-03-09 of Marathon Oil Corporation). For the stock of Apple Inc we establish limits of log(0.48) and log(1.33), consistent with its historical extrema. These truncation limits (r_{min}) will also be used when we perform numerical integration to calculate the Expected Shortfall of probability density functions (this is using eq. (3) below with integration limit r_{min} instead of $-\infty$).

III. B. How the Expected Shortfall from Historical Simulation depends on the size of the dataset

The Expected Shortfall of a return r considered as a continuous random variable can be defined as follows:

$$\mathrm{ES} \;=\; \frac{-1}{\alpha} \int_{-\infty}^{r_0} r\,\mathrm{pdf}(r)\,dr \quad , \tag{3}$$

where $\mathrm{pdf}(r)$ is the probability density function of the return; r_0 is the value of the return such that $\int_{-\infty}^{r_0} \mathrm{pdf}(r)\,dr \;=\; \alpha$, being the *confidence level* $\alpha < 1$ the number which represents the total probability of the most adverse returns to be considered in the calculation of the ES. The $(-1/\alpha)$ multiplicative factor is sometimes omitted from the definition of the ES, see for instance eq. (8.52) of (McNeil et al., 2005). Other equivalent definitions exist, e.g. $\mathrm{ES} \;=\; \mathbb{E}[\,r\,|\,r \leq -\mathrm{VaR}\,]$ where $\mathbb{E}$ indicates expected value, and ES, VaR are the Expected Shortfall and Value-at-Risk for confidence level α of the analyzed return.

Despite definition (3), the Expected Shortfall is frequently calculated by considering just observed values of a given return r_i. For example, the European Banking Authority (European Banking Authority, 2020) specifies the formula that follows for its calculation:

$$\mathrm{ES} \;=\; \frac{-1}{\alpha\,N_p} \left[\left(\sum_{i=1}^{\lfloor \alpha\,N_p \rfloor} r_i \right) + \left(\,\alpha\,N_p - \lfloor \alpha\,N_p \rfloor\,\right) r_{\lfloor \alpha\,N_p \rfloor + 1} \right] \quad , \tag{4}$$

where the $\lfloor x \rfloor$ signs indicate the integer part (floor) of x and the indices (i) of r_i are ordered so that r_i are monotonically increasing; we set $\alpha = 0.025$ (2.5 %). This way of proceeding ignores altogether other non-observed values of the returns. Such values were possible, but do not take part in the calculation; hence, potentially important information is discarded. On the one hand, eq. (4) has the advantage that it does not require any assumption on the actual distribution of the returns (i.e. if it is normal, generalized hyperbolic, etc.). On the other hand, the cost of making this assumption may be to distort the calculated Expected Shortfall strongly. The arbitrariness of the choice of the underlying probability distribution can be partly mitigated by analyzing different ones, as we do in this research work.

To quantify whether the usage of Historical Simulation (eq. (4)) severely distorts the calculation of the ES, we proceed as follows. We fit the collection of returns of a given financial product for a given time range (see Tab. 1) to four different probability density functions. Among them, we choose the one whose loss function (eq. (2)) is minimal, i.e., which has the maximum likelihood (see bold numbers in Tab. 2). We then generate synthetic datasets using the parameters of the chosen distribution. Each synthetic dataset consists of N_{size} points. For each value of N_{size} we generate half a million of synthetic datasets; for each of them, we calculate the ES using eq. (4). We then calculate the mean, median, standard deviation and 95% confidence interval of this collection of 500k ES's for each N_{size}. For the returns of the analyzed BASF bond, these quantities are presented in Fig. 3-top. The solid lines of the upper subplots clearly indicate that the Expected Shortfall tends to increase, in a monotonical manner, with the sample size. It converges to steady values (plateau) for high values of N_{size}, but requires relatively high numbers of returns to approach it. For example, the mean and median of the ES of the synthetic data are far from their converged values for $N_{size} < 100$.

Such a monotonical increase is a consequence of the highly nonlinear definition of the Expected Shortfall. For example, if we calculate the average of the synthetic data instead of the ES, we will notice that there is not a trend of it with the size of the dataset. This can be viewed in Fig. 3-bottom, which presents the mean, median, standard deviation, and confidence interval of the *average* of each synthetic dataset that was used in the calculations for Fig. 3-top. Here the solid lines are horizontal; the sizes of the differences of the mean of the averages for different values of N_{size} are far lower than the standard deviation (orange curve in Fig. 3-top, left).

In Fig. 2 we present a similar analysis to that of Fig. 3. The outputted results correspond to synthetic time series generated from observed logarithmic returns of the Apple stock. Further analogous results for the other analyzed financial products are presented in the Supporting Information. They all support the thesis that the average ES monotonically increases with the sample size until it reaches a plateau.

III. C. Expected Shortfall from fitting to small datasets

The results presented in the previous section indicate that the usage of Historical Simulation (i.e., using observed values only) in the calculation of the ES is potentially inaccurate and prone to systematic errors (underestimation of the ES). This drawback can be mitigated by increasing the sample size. However, such a solution may not always be either feasible or accurate. Simply increasing the number of returns used in the HS calculation would probably require using older data, which may be stale. The economic conditions, as well as the inner operation of a company, tend to change over time; hence the obsolescence of data may lead to inaccurate results. Moreover, there exist illiquid products, like many corporate bonds, for which the returns are unknown for many dates. In those cases, one is forced to perform calculations with few returns (datasets of small size), thus leading to potentially severe inaccuracies, as indicated by Figs. 3 and 2 (which correspond to absolute returns of a bond price and logarithmic returns of a stock price, respectively). Can this inconvenience be overcome? In this section, we present a method to mitigate the problem.

Since the inaccuracies in the Expected Shortfall calculation are a consequence of not considering non-observed, though possible, extreme values, the impact of their neglect can be eased by finding an approximation to them. We can infer the probability density function from the analyzed dataset, as indicated above. Even if this dataset consists of a few points, the fitting will provide probabilities for extreme values, which can be used in the calculation of the Expected Shortfall.

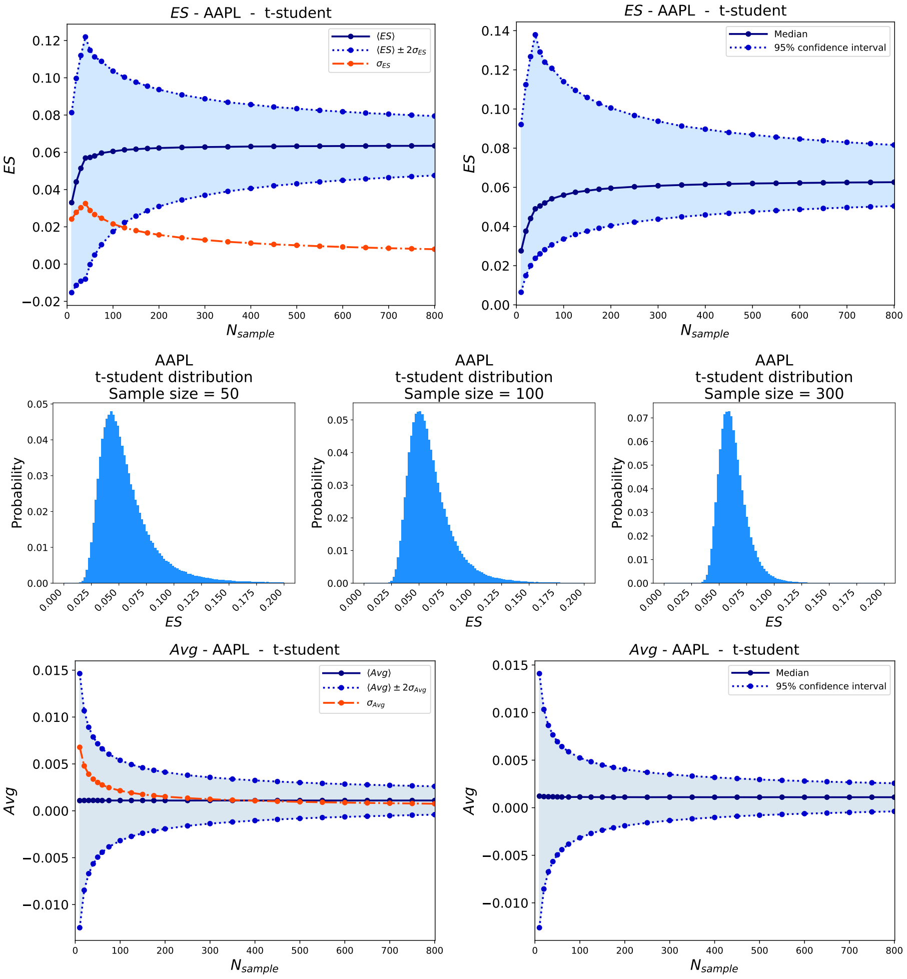

Figure 2: Top: Relationship between the ES calculated using the historical method (HS) and the number of data points of the sample size. Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval.

Center: Histograms of the expected shortfalls obtained with synthetic data of the absolute returns of a stock (Apple) as a function of the number s of the generated random values for each ES calculation. Left: $s = 50$; Center: $s = 100$; Right: $s = 300$.

Bottom: Relationship between the average of the synthetic data (identical to those of the top figures) and the number of data points of the sample size. Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval. The synthetic data were generated using a non-centered t-student distribution.

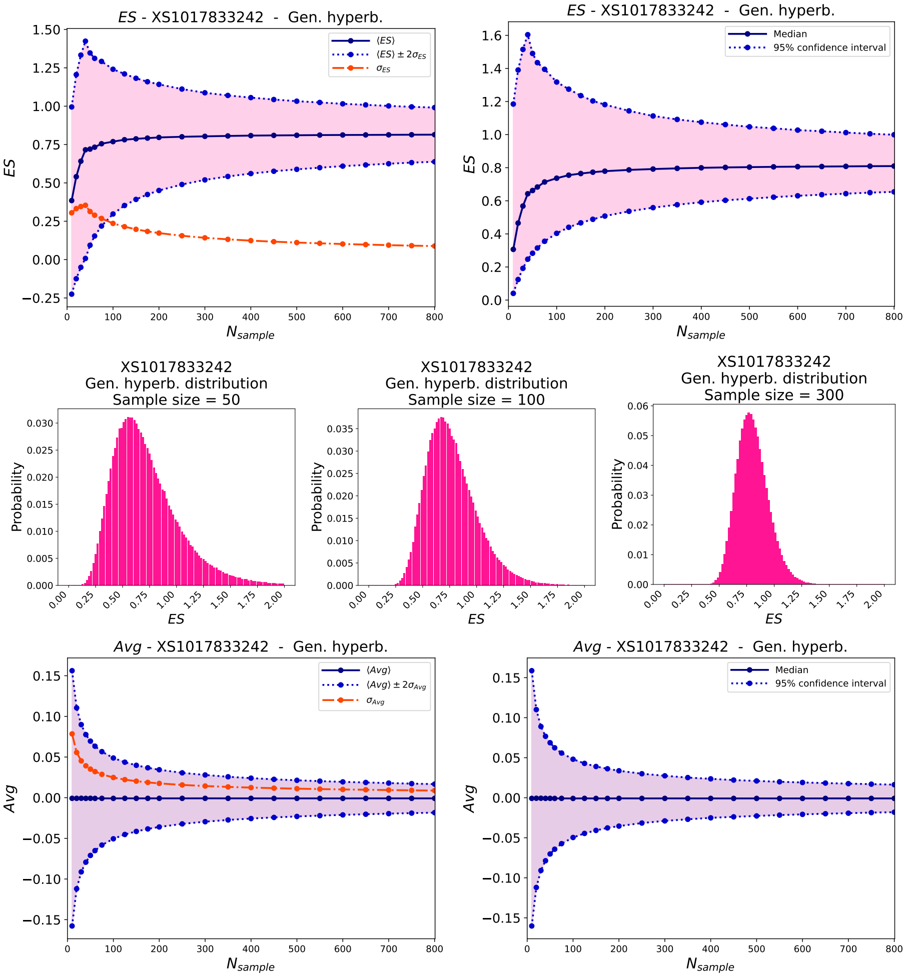

Figure 3: Top: Relationship between the ES calculated using the historical method (HS) and the number of data points of the sample size. Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval.
Center: Histograms of the expected shortfalls obtained with synthetic data of the absolute returns of a bond (ISIN XS1017833242) as a function of the number s of the generated random values for each ES calculation. Left: $s = 50$; Center: $s = 100$; Right: $s = 300$.
Bottom: Relationship between the average of the synthetic data (identical to those of the top figures) and the number of data points of the sample size. Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval. The synthetic data generated using a generalized hyperbolic distribution.

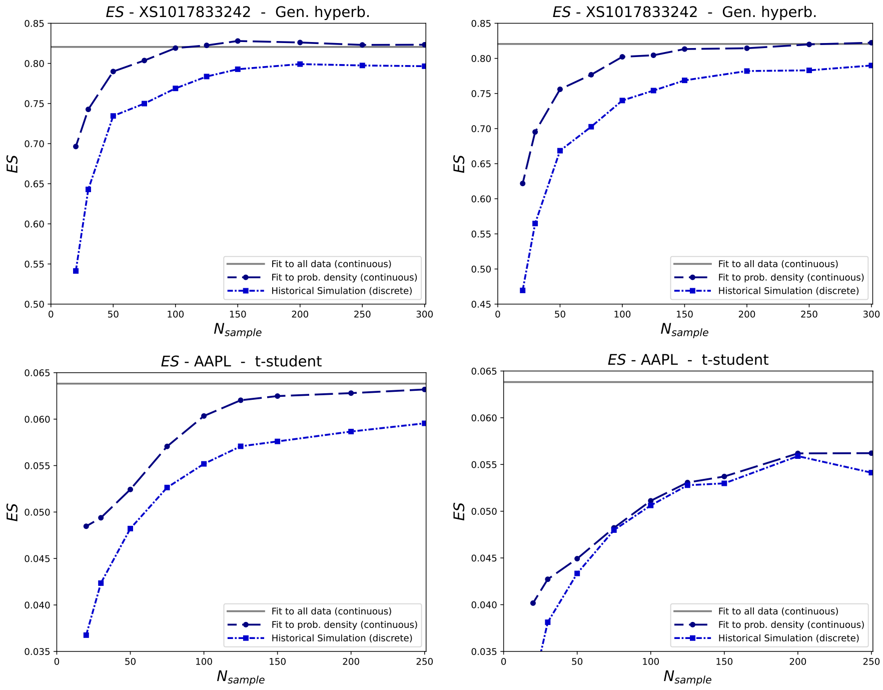

Figure 4: Comparison of the mean (left) and median (right) of the Expected Shortfall of datasets of different sizes (N_{sample}). The dash-dotted curves (Historical Simulation) correspond to ES of collections of observed returns calculated with eq. (4); the dashed curves correspond to ES calculated from fitting those collections of observed returns to fat-tailed distributions. Gray horizontal lines correspond to the ES from the fitted distribution of the whole dataset. The plots on top correspond to a BASF bond (fitted to generalized hyperbolic distributions); the plots on the bottom correspond to the AAPL stock (fitted to non-centered t-student distributions).

We exemplify this way to proceed with the results displayed in Fig. 4. In the plots on top, which correspond to the BASF bond, the dash-dotted lines correspond to the mean and median of the 2000 values of the ES of different sets of synthetic data using HS (from eq. (4)). The dashed lines represent the mean and median of the ES obtained by fitting each dataset to a generalized hyperbolic probability density function, and then using such continuous functions to calculate the ES. Since the continuous probability density functions are less prone to underestimate the extreme values (*tails*) than Historical Simulation, the values of the dashed line are always above the values of the dash-dotted line. These results indicate that, if we take the ES of the fitting to the large (whole) observed dataset (gray horizontal line in Fig. 4-top) as our baseline, then fitting to probability distributions provides more accurate results than HS. The effect is especially strong for low values of the sample size.

The analyzed BASF bond can be considered a proxy of other corporate bonds, which can be illiquid. For illiquid bonds, just a few returns are known, and hence taking values of a proxy (BASF) bond is expected to reasonably give an account of the modeled product (illiquid bond). However, this is not the case for stocks (shares of companies traded in stock exchanges). Though there exist many small companies whose stocks have very low trading volume (and may not be purchased even once a day), the vast majority of the volume of traded stocks corresponds to very liquid products, which are traded numerous times a day. Therefore, the fitting of observed data for calculating accurate ES may, in principle, not look necessary for stocks, because their daily prices are known. Conversely, we think that the fitting procedure is also worthwhile for stocks. This is because the number of days to be used for calculating the ES (N_r in eq. (4)) is arbitrary. If one wants to calculate risks with a given horizon (e.g. 126 days, the usual number of trading days in a semester), the market conditions may have changed during that time, making the oldest data stale. In that case, it may be more appropriate to choose e.g. 63 days instead. But from our previous analysis (Fig. 2-top) we know that such a small number of data would distort ES if calculated through historical simulation, leading to a non-converged value (below the plateau). Therefore, a wiser way to proceed would be to use recent data, to fit them to a fat-tailed distribution, and finally to calculate the ES of that distribution.

In Fig. 4-bottom we present data analogous to those of Fig. 4-top, yet with the corresponding calculations performed differently. For every sample, with size N_{sample} (x-axis of the figures), we no longer randomly generated N_{sample} returns of the price of the analyzed product. In contrast, we took the last N_{sample} returns of the stock price. This was done for every single trading date of the analyzed time interval (see Tab. 1). Therefore each point displayed in Fig. 4-bottom is not the average of 2000 trials, but the average of a number of trials equal to the total number of returns of the time interval (5 years) minus $(N_{sample} - 1)$.

In Fig. 4-bottom we also present a horizontal gray line which indicates the ES from the distribution fitted to the whole dataset. Note that this is just a reference; it does not need to be equal to the calculated values (blue lines). This is because the ES may abruptly change for sudden strong price drops. For example, if $N_{sample} \simeq 262$ the ES may be 0.05 the first year, 0.01 the second year, 0.03 the third year, etc., and the average of these numbers would not need to be the ES of the whole (5-year) dataset, which may closer to 0.05 if the strongest price dips concentrated in that period.

Fitting a dataset to probability density functions is more computationally demanding than a simple calculation of the ES using Historical Simulation. However, due to the capabilities of present-day computing facilities (García-Risueño and Ibáñez, 2012), such calculations are affordable. The calculations whose results are presented in this article were performed on a personal laptop (MacBook Pro 13-inch M1 2020, i.e. not on a cluster or other supercomputing facility), and were carried out within a few weeks. The code was run in `Python 3.11`, using recent versions of `numpy` and `scipy.stats` modules (Harris et al., 2020; Virtanen et al., 2020).

IV. Conclusion

In this article we have shown a presentation on how the method of Historical Simulation tends to systematically underestimate the Expected Shortfall of random variables like the returns of financial products. This is due to the fact that observed sample datasets are forcedly finite, and hence cannot cover the virtually infinite possible values which correspond to continuous probability distributions. This limitation has a low impact when calculating some quantities like the sample average. Yet it has a dramatic impact on the calculation of the ES, which strongly depends on the minimum values of the sample, which in turn tend to be lower for higher numbers of points in the sample. We have indicated two methods to mitigate the inaccuracies in the ES: increasing the sample size and, –most recommended– fitting the sample dataset to a continuous, fat-tailed probability density function to be used in the calculation of the Expected Shortfall. We expect that our research work provides insights on important statistical properties to Finance practitioners, and that it helps in their risk management activities.

Acknowledgments

The author is grateful to Prof. Eduardo Ortas and to Prof. Jesús Asín (University of Zaragoza) for interesting discussion and revision of the manuscript.

References

Barzilai, J., Borwein, J. M., 1988. Two-point step size gradient methods. IMA journal of numerical analysis 8 (1), 141–148.

Carrillo Menéndez, S., Hassani, B. K., 2021. Expected Shortfall reliability-added value of traditional statistics and advanced artificial intelligence for market risk measurement purposes. Mathematics 9 (17), 2142.

Choudhry, M., 2013. An introduction to Value-at-Risk. John Wiley & Sons.

European Banking Authority, 2020. Final Draft RTS on the calculation of the stress scenario risk measure under Article 325bk(3) of Regulation (EU) No 575/2013 (Capital Requirements Regulation 2 – CRR2). EBA publications.

Garcia-Risueño, P., 2023. https://github.com/pablogr/FIN_paper_ES_fat_tails.

García-Risueño, P., Ibáñez, P. E., 2012. A review of high performance computing foundations for scientists. International Journal of Modern Physics C (IJMPC) 23, 1230001.

Garcia-Risueño, P., Ortas, E., Moneva, J. M., 2023. The effect of fat tails on rules for optimal pairs trading. submitted.

Hafez, P., Lautizi, F., 2019. Machine learning and event detection for trading energy futures. In: Guida, T. (Ed.), Big Data and Machine Learning in Quantitative Investment. John Wiley & Sons., pp. 169–185.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T. E., Sep. 2020. Array programming with NumPy. Nature 585 (7825), 357–362.

Johnson, N. L., Kotz, S., Balakrishnan, N., 1995. Continuous univariate distributions, volume 2. Vol. 289. John wiley & sons.

Liu, M., Staum, J., 2010. Stochastic kriging for efficient nested simulation of Expected Shortfall. Journal of Risk 12 (3), 3.

López de Prado, M., 2018. Advances in financial machine learning. John Wiley & Sons.

Maio, S., Maurette, M., et al., 2017. Risk measurement using extreme values theory. Proceedings of the 2017 MACI.

McNeil, A. J., Frey, R., Embrechts, P., 2005. Quantitative risk management: concepts, techniques and tools. Princeton University Press.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 17, 261–272.

Wong, W. K., 2008. Backtesting trading risk of commercial banks using Expected Shortfall. Journal of Banking & Finance 32 (7), 1404–1415.

Supporting information of the article *Historical Simulation systematically underestimates the Expected Shortfall*

December 2, 2024

Pablo García-Risueño

risueno@unizar.es, garcia.risueno@gmail.com

Abstract

In this document, we present information to complement the main article, which was not included for brevity's sake. Graphs for all four analyzed financial products are displayed. The presented figures include those describing the fitting of distributions to the observed datasets and plots of the quantiles of data from distributions vs quantiles from HS datasets. We also display the relationship between the ES and the sample size, and between the mean and the sample size, as well as comparisons of the ES from datasets from HS and from fittings to such datasets. The names of the sections of this document are equal to those of the main article, with each section complementing the homonym section of the main paper.

JEL codes: C22, C54, C61, C87.

Fitting to probability distributions

For the sake of brevity, in the main article we only displayed figures of the fitting for one of the analyzed financial products. In Figs. 1, 2 and 3 we display the fittings corresponding to the other three products.

To provide the reader with a deeper insight on the tails of these histograms we present Figs. 4 to 7. They display Q-Q probability plots of the quantiles of synthetic data generated vs. data from Historical Simulation. The shown quantiles are 0.01, 0.02, 0.03, 0.04, 0.05, 0.10, 0.15, . . ., 0.85, 0.90, 0.95, 0.96, 0.97, 0.98, 0.99. Each x axis represents the quantiles of sets of returns of sizes N=100, 300 and 1000 consisting of the N returns till the latest date displayed in Tab. 1 of the main article (logarithmic returns for stocks, absolute returns for bonds). Since these are observed values, the x axis represents the quantiles from Historical Simulation. On the other hand, each y axis represents the quantiles of large arrays of synthetic data generated using the fitting parameters displayed in Tab. 2 of the main article. These synthetic data were truncated to avoid too extreme values as explained in the main article. Since these arrays are large (10 million of datapoints) they are expected to be good proxies of the actual continuous probability distributions. Note that since we only took *one* HS data array for each size, the results displayed in Figs. 4 to 7 are mainly illustrative, i.e. they do not provide a comprehensive of the features of the data.

In a Q-Q plot, the different behavior of the distributions underlying the data of the two displayed axes manifests as a relatively large vertical distance of the plotted dots from the diagonal ($x = y$) line. In Figs. 4 to 7 one notices that, generally speaking, this distance lowers for higher sizes of the HS dataset (N). This confirms that larger datasets tend to give a more accurate account of the extreme values of the analyzed random variables, which corresponds to calculations of the Expected Shortfall with a higher accuracy. In addition, Figs. 4 to 7 show that fitting to a normal distribution provides estimates of the quantiles, which are much less accurate than those from fat-tailed distributions. As expected from the values of the loss function displayed in Tab. 2 of the main article, the best matchings are obtained for the non-centered t-student distribution and the generalized hyperbolic distribution.

How the Expected Shortfall from Historical Simulation depends on the size of the dataset

Also for the sake of brevity, in the main article we displayed the relationship of the mean and median of the ES and the average of the returns for only one distribution (generalized hyperbolic) of one product (the BASF bond). In Figs. 8 to 11 we present the corresponding plots for all four distributions for that product; in Figs. 12 to 15 we present the corresponding mean and median of the averages of the synthetic datasets. The synthetic time series used in Figs. 8 and 12 are the same, which also holds for the other distributions (Figs. 9 and 13, 10 and 14, as well as 11 and 15). For all, we also present example histograms for given sizes of the dataset (N_{size} equal to 50, 100 and 300). The histograms of the averages have an approximately Gaussian shape, as one could forecast using the Law of large numbers. Conversely, the histograms of the Expected Shortfall look very skewed in all cases. For both the ES and the average, higher values of N_{size} lead to narrower histograms.

In Figs. 16, 17 and 18 we also present the expected shortfall and averages for different sample sizes for three further financial products: a bond of Charles Schwab and the stocks of Apple and Shell p.l.c., respectively. The observed data correspond to the dates presented in Tab. 1 of the main article, and the synthetic distributions were generated with the fitted parameters displayed in Tab. 2 of the main article.

Our results indicate that the choice of the fat-tailed distribution to be used for fitting the observed data has a strong impact on the ES of the fitted distribution. To solve this drawback, we suggest using one of the two following procedures: i) Consider only one distribution kind (e.g. generalized hyperbolic) for all the products of a given kind (e.g. for all the bonds of your portfolio); ii) Choose the distribution whose loss function from maximum likelihood is lowest. The latter is the approach that we followed in this article. The case of the Apple stock hints that if the values of the loss functions of two different distributions are similar, then the values of the ES will also be alike. For this case, as it can be viewed in Tab. 2 of the main article, the loss function for the cases of non-centered t-student and generalized hyperbolic distributions are -2.522572 and -2.522517, respectively, which is a relatively small difference. In Fig. 19 we present the averages of mean, median and standard deviation of sets of synthetic data of different sizes for both distributions. The synthetic data were generated with the fitted parameters displayed in Tab. 2 of the main article. We observe that these values are reasonably similar.

Expected Shortfall from fitting to small datasets

In Figs. 20 to 29 we give a broader scope to the results presented in the corresponding section of the main article.

The displayed data were calculated as explained in the main article, with 2000 simulations for each dot. The exception is the Charles Schwab bond, where 4000 simulations were used in the range $N = 20, \ldots, 125$, 2000 simulations were used for $N = 150, 200$ and 1000 simulations were used for $N = 250$. Fig. 20 represents the mean and median of the ES calculated using Historical simulation for synthetic datasets using a normal distribution (dash-dotted lines) and using fitting to probability density functions for each synthetic dataset (being the ES calculated from the probability density function using numerical integration). In Fig. 21 we enhance the information presented in Fig. 20, including standard deviation and confidence intervals. Figs. 22 and 23 represent the same quantities as Figs. 20 and 21, except for the fact that the distributions are not normal, but generalized hyperbolic. In Fig. 22-left it can result unexpected to find some values of the ES above the gray line. We attribute this to the fact that the mean is sensitive to outliers and each point of the curves was calculated from only 2000 values of the ES due to computational limitations. The medians (Fig.

22-right) do not present the referred unexpected feature.

In Fig. 24-left we observe that the mean ES from historical simulation becomes higher than that from fitting beyond the lowest sample sizes. This is probably due to the fact that the distribution is prone to outliers, as can be noticed by comparing means and medians (24-right). The fact that the ES from fitting evolves more smoothly hints that the fitting might make the results more stable, less prone to outliers.

Figures

Below we present the figures referenced in the sections above.

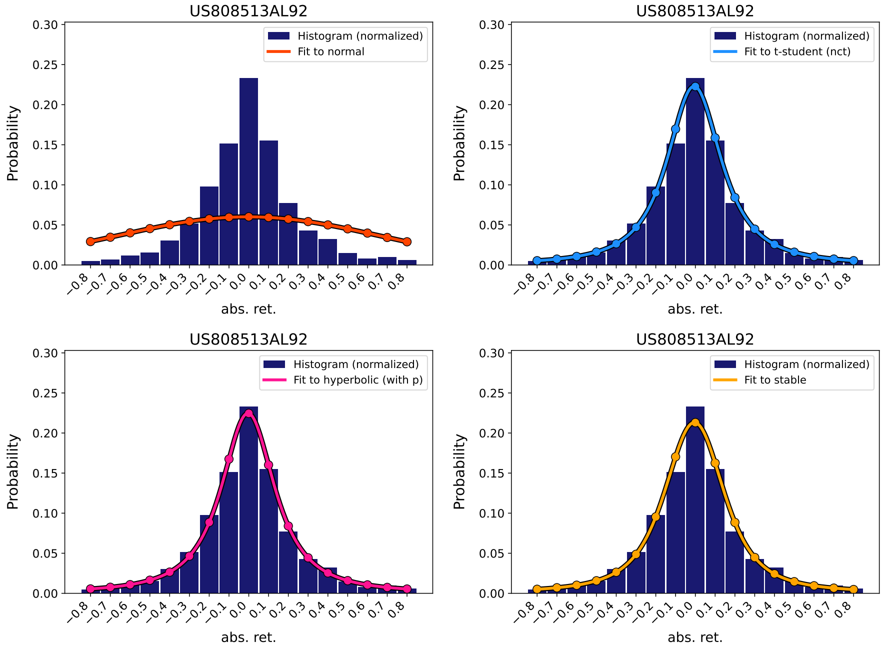

Figure 1: Fitting of absolute returns of the price of a bond (ISIN US808513AL92, from Charles Schwab) to different probability density functions. Top, left: Normal; Top, right: Non-centered t-student; Bottom, left: Generalized hyperbolic; Bottom, right: Lévy stable.

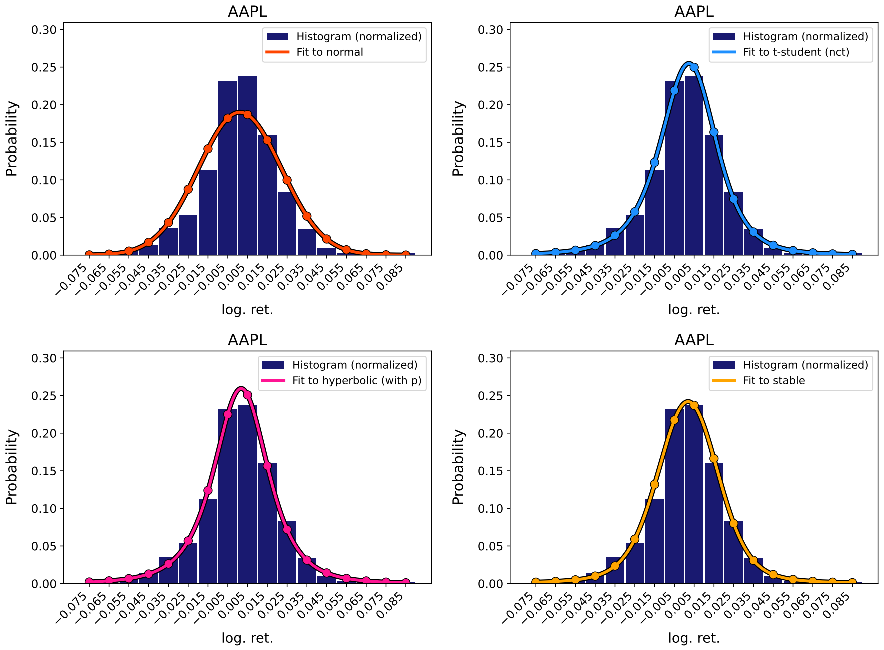

Figure 2: Fitting of absolute returns of the price of a stock (Apple Inc.) to different probability density functions. Top, left: Normal; Top, right: Non-centered t-student; Bottom, left: Generalized hyperbolic; Bottom, right: Lévy stable.

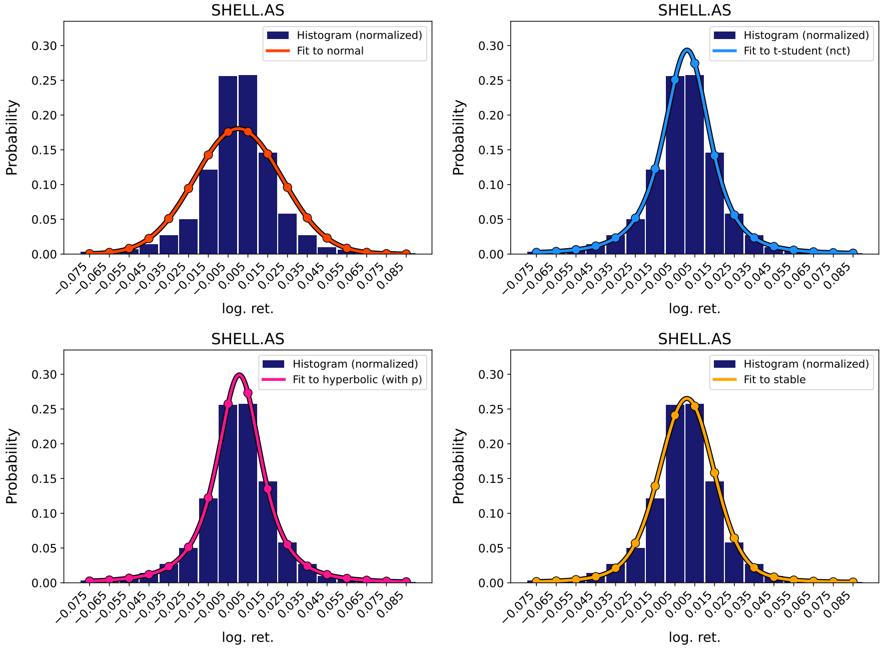

Figure 3: Fitting of absolute returns of the price of a stock (Shell p.l.c.) to different probability density functions. Top, left: Normal; Top, right: Non-centered t-student; Bottom, left: Generalized hyperbolic; Bottom, right: Lévy stable.

Declaration of interest statement

There is no conflict of interest to declare.

Author information

E-mail address: risueno@unizar.es, garcia.risueno@gmail.com
ORCID: 0000-0002-8142-9196.

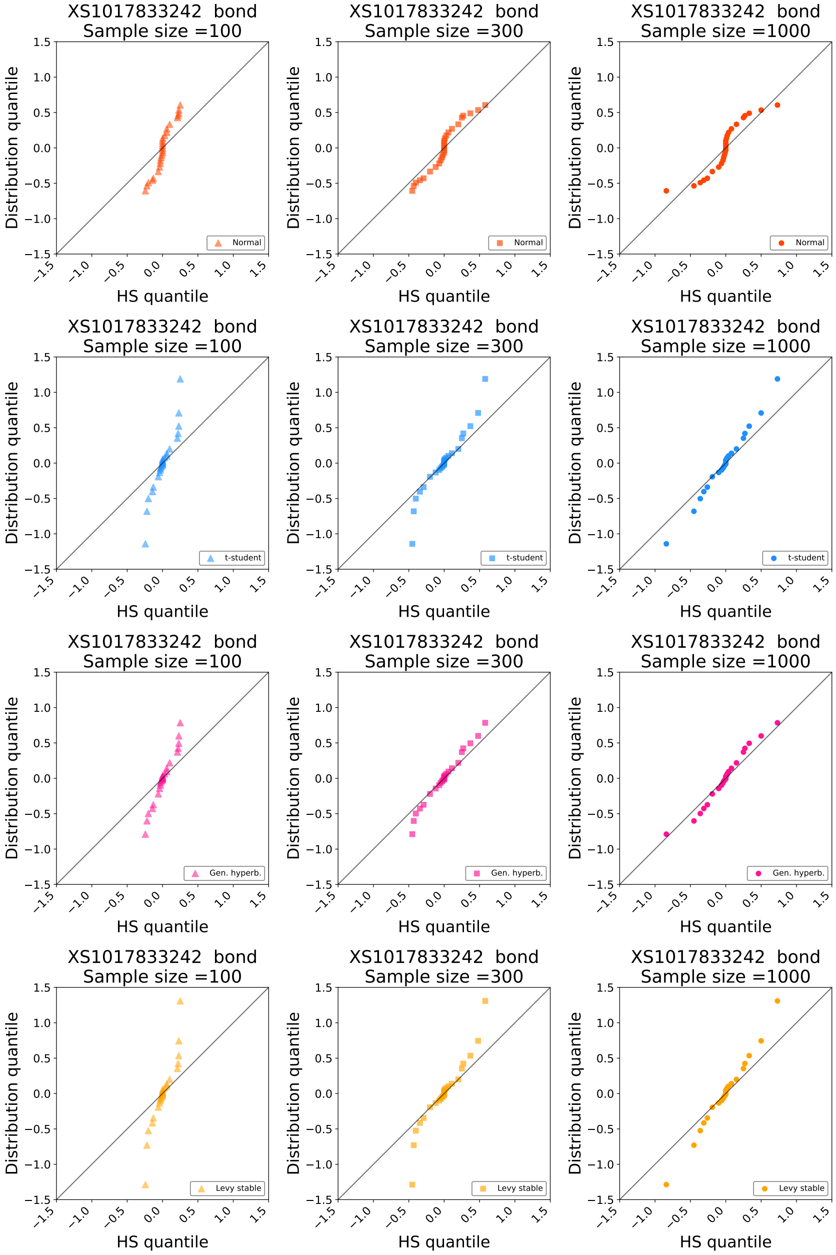

Figure 4: Plots of quantiles of a large array of data generated from the fitted distributions to returns of the bond with ISIN XS1017833242 (y-axis) vs. quantiles of its N (=100, 300, 1000) last returns (x-axis).

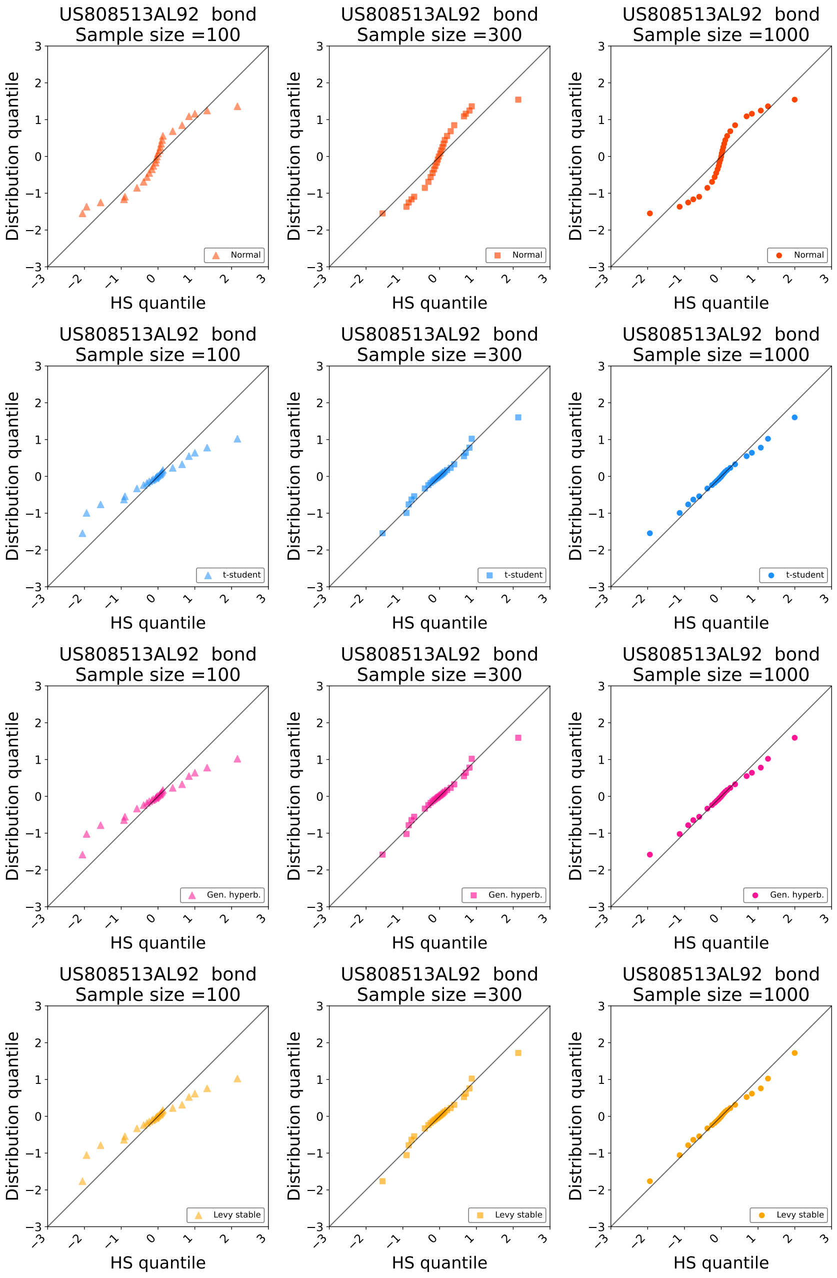

Figure 5: Plots of quantiles of a large array of data generated from the fitted distributions to returns of the bond with ISIN US808513AL92 (y-axis) vs. quantiles of its N (=100, 300, 1000) last returns (x-axis).

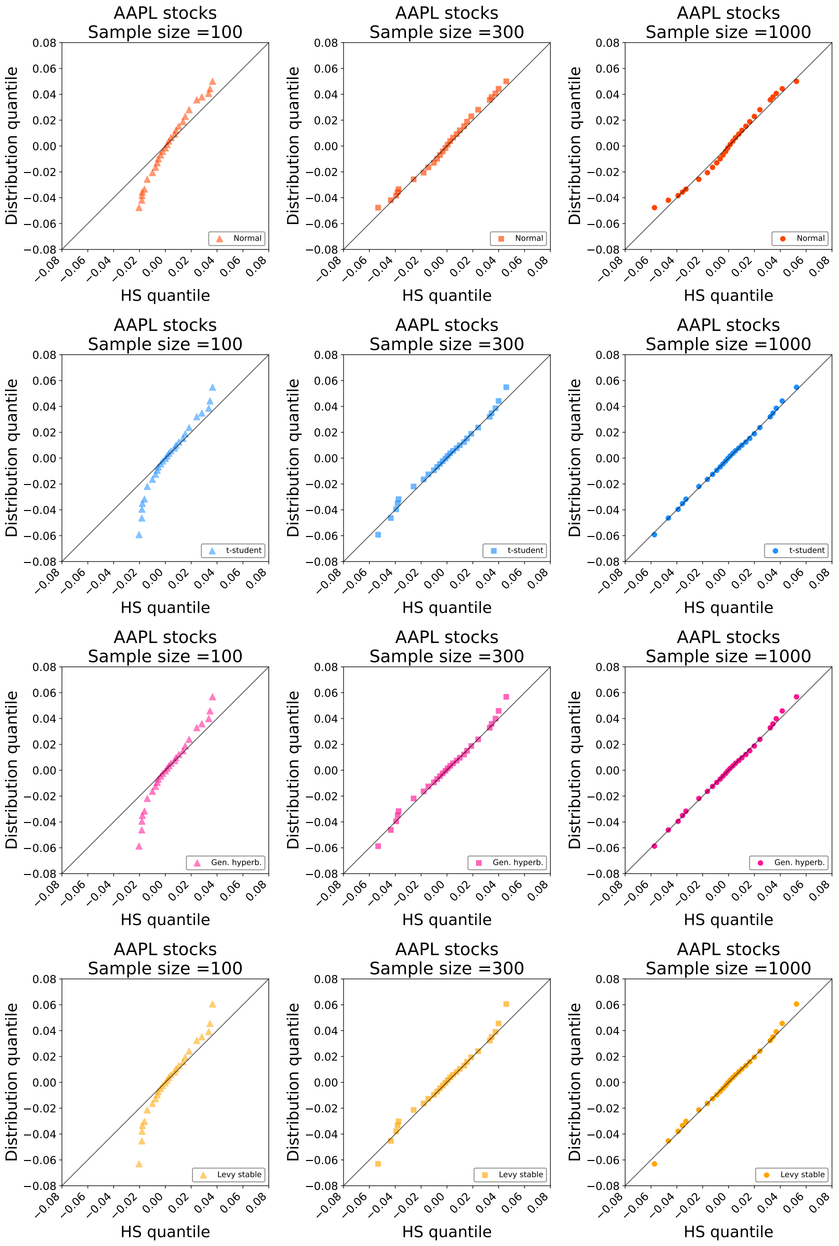

Figure 6: Plots of quantiles of a large array of data generated from the fitted distributions to returns of the Apple stock (y-axis) vs. quantiles of its N (=100, 300, 1000) last returns (x-axis).

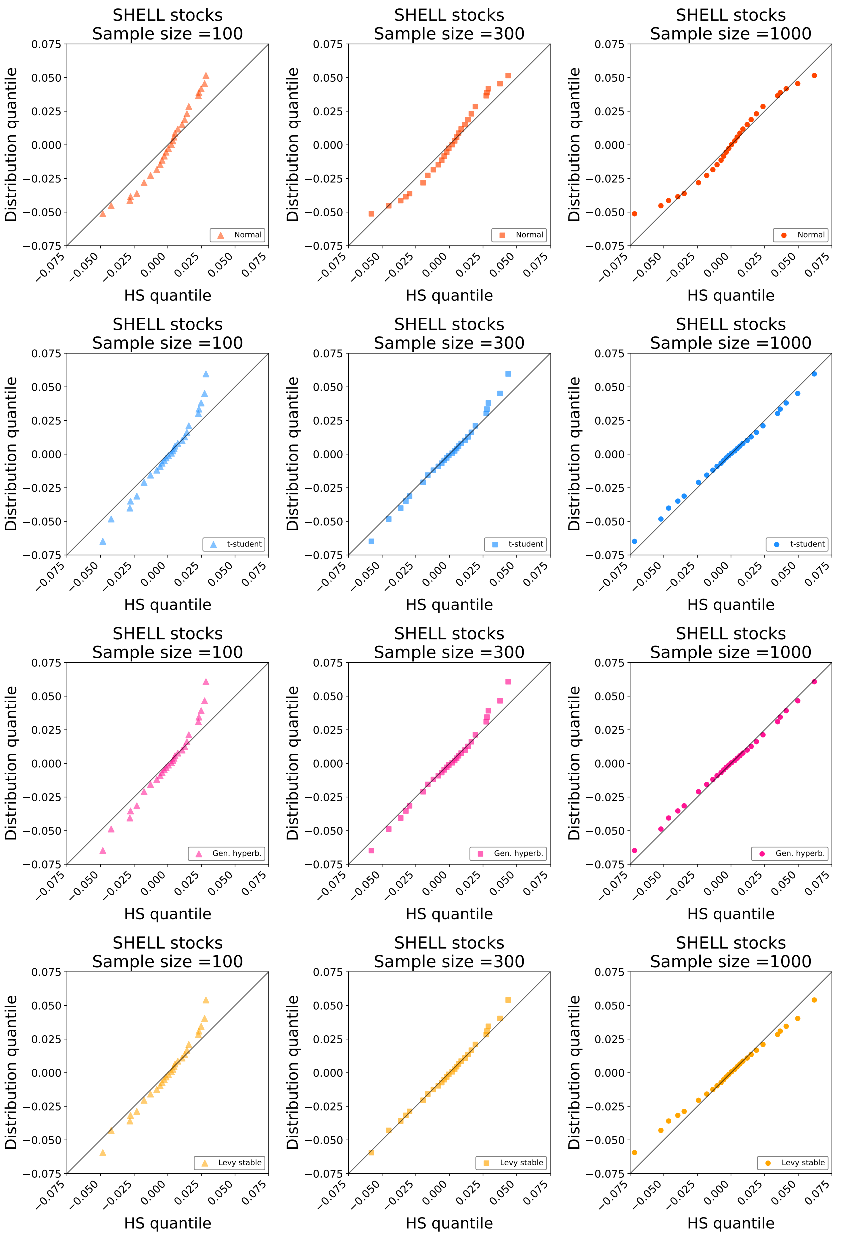

Figure 7: Plots of quantiles of a large array of data generated from the fitted distributions to returns of the Shell p.l.c. stock (y-axis) vs. quantiles of its N (=100, 300, 1000) last returns (x-axis).

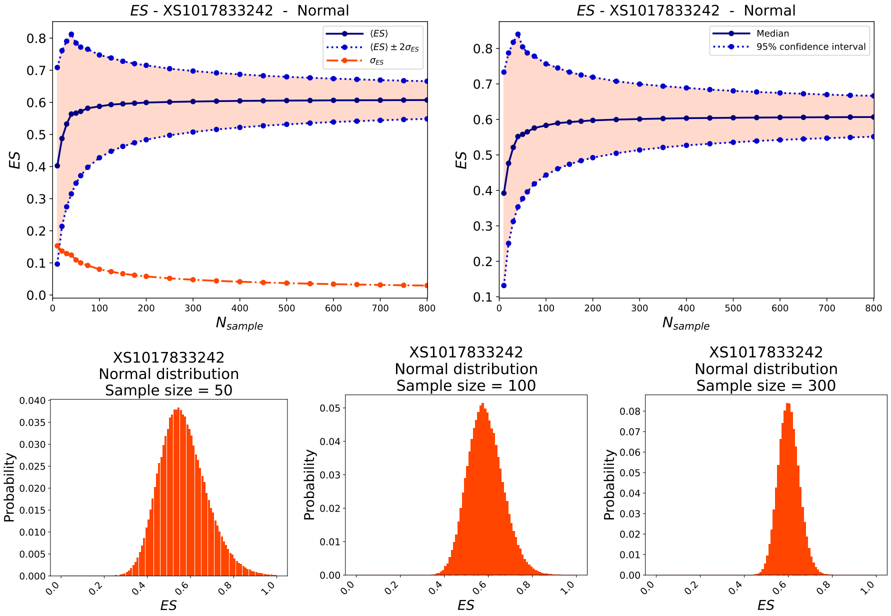

Figure 8: Relationship between the expected shortfall of the synthtetic data and the number of data points of the sample size. Top: Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval. Bottom: Histograms for three different sample sizes. Synthetic data from a normal distribution, whose parameters were fitted to historical data of the absolute returns of a bond (ISIN XS1017833242).

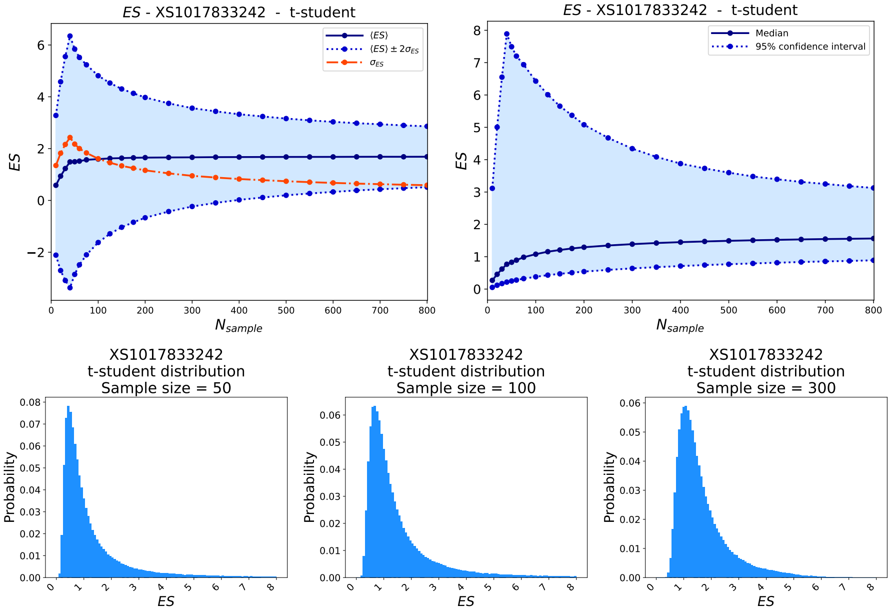

Figure 9: Relationship between the expected shortfall of the synthtetic data and the number of data points of the sample size. Top: Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval. Bottom: Histograms for three different sample sizes. Synthetic data from a non-centered t-student distribution, whose parameters were fitted to historical data of the absolute returns of a bond (ISIN XS1017833242).

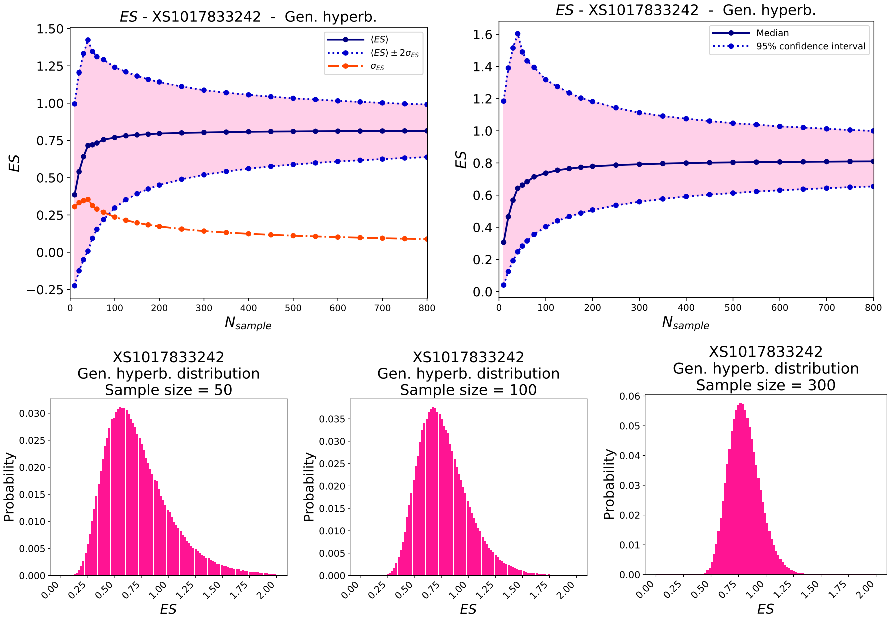

Figure 10: Relationship between the expected shortfall of the synthtetic data and the number of data points of the sample size. Top: Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval. Bottom: Histograms for three different sample sizes. Synthetic data from a generalized hyperbolic distribution, whose parameters were fitted to historical data of the absolute returns of a bond (ISIN XS1017833242).

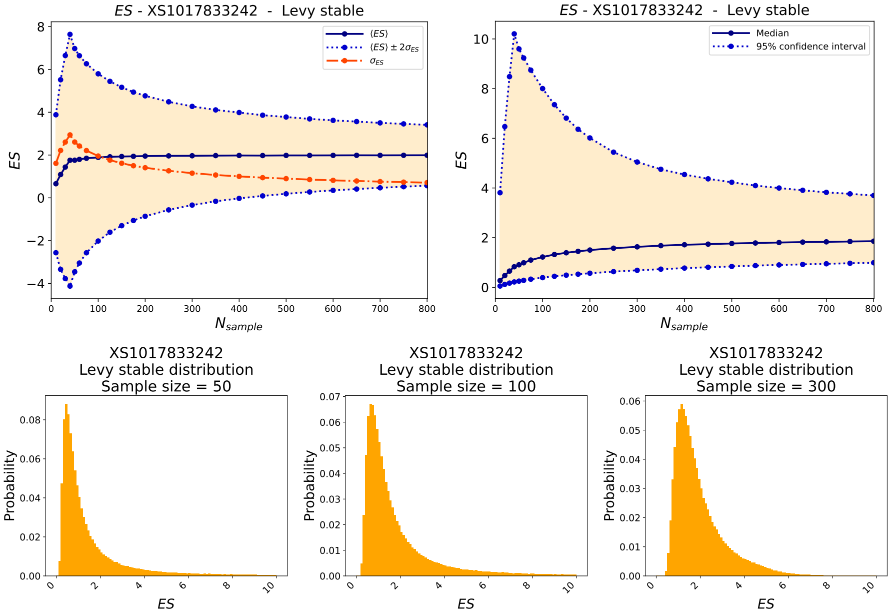

Figure 11: Relationship between the expected shortfall of the synthtetic data and the number of data points of the sample size. Top: Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval. Bottom: Histograms for three different sample sizes. Synthetic data from a Lévy stable distribution, whose parameters were fitted to historical data of the absolute returns of a bond (ISIN XS1017833242).

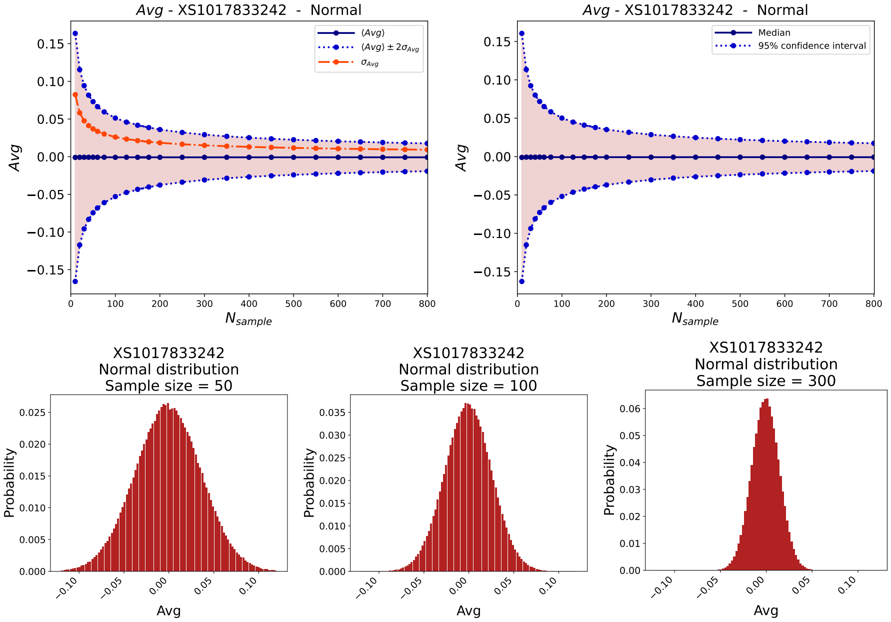

Figure 12: Relationship between the average of the synthtetic data and the number of data points of the sample size. Top: Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval. Bottom: Histograms for three different sample sizes. Synthetic data from a normal distribution, whose parameters were fitted to historical data of the absolute returns of a bond (ISIN XS1017833242).

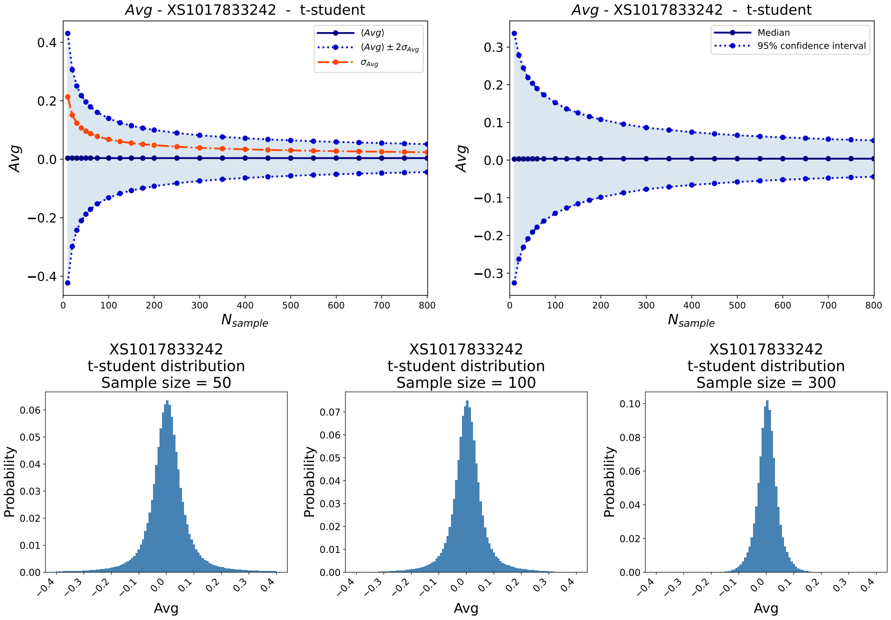

Figure 13: Relationship between the average of the synthtetic data and the number of data points of the sample size. Top: Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval. Bottom: Histograms for three different sample sizes. Synthetic data from a non-centered t-student distribution, whose parameters were fitted to historical data of the absolute returns of a bond (ISIN XS1017833242).

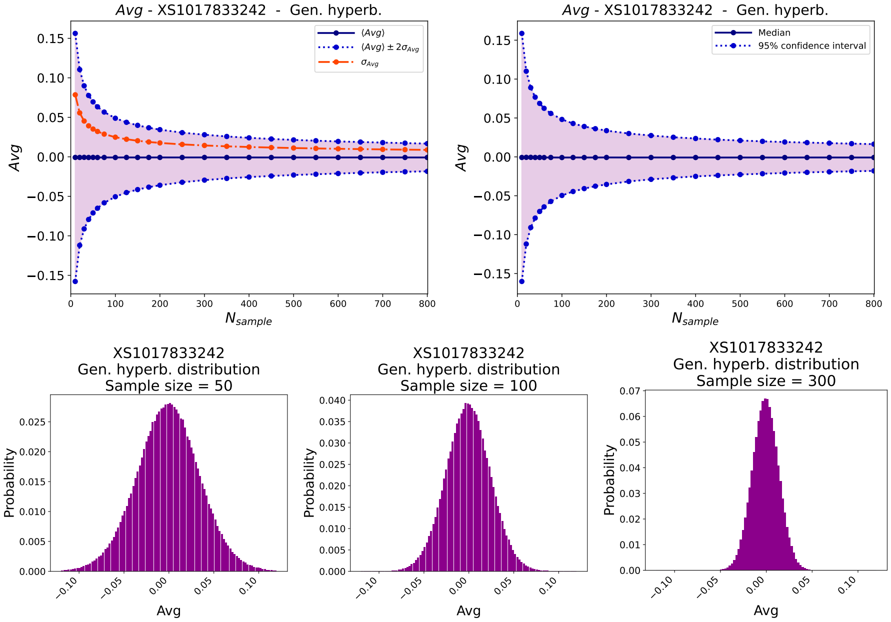

Figure 14: Relationship between the average of the synthtetic data and the number of data points of the sample size. Top: Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval. Bottom: Histograms for three different sample sizes. Synthetic data from a generalized hyperbolic distribution, whose parameters were fitted to historical data of the absolute returns of a bond (ISIN XS1017833242).

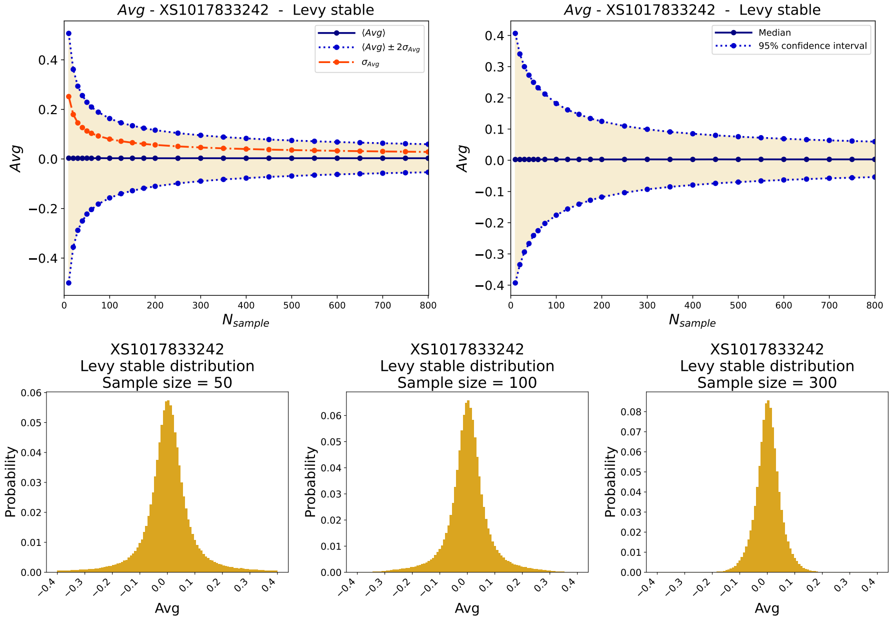

Figure 15: Relationship between the average of the synthtetic data and the number of data points of the sample size. Top: Left: Standard deviation (red) and average plus/minus twice the standard deviation (blue); Right: Median and 95% confidence interval. Bottom: Histograms for three different sample sizes. Synthetic data from a Lévy stable distribution, whose parameters were fitted to historical data of the absolute returns of a bond (ISIN XS1017833242).

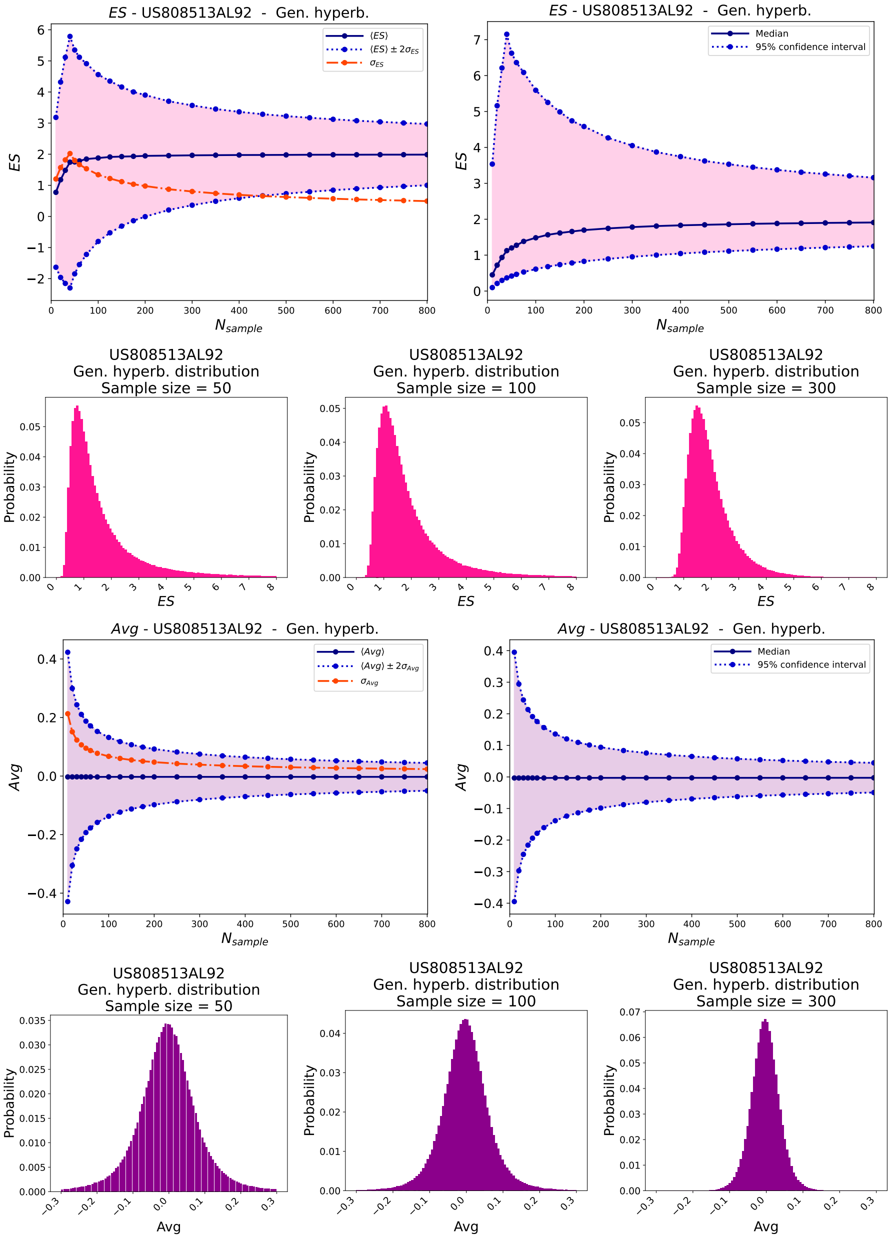

Figure 16: Relationship between the expected shortfall (row 1) and the average (rows 3) of the synthtetic data and the number of data points of the sample size. The left plots present the standard deviation (red) and average plus/minus twice the standard deviation (blue); the right plots present the median and 95% confidence interval. Rows 2 and 4 present corresponding histograms for three different sample sizes.

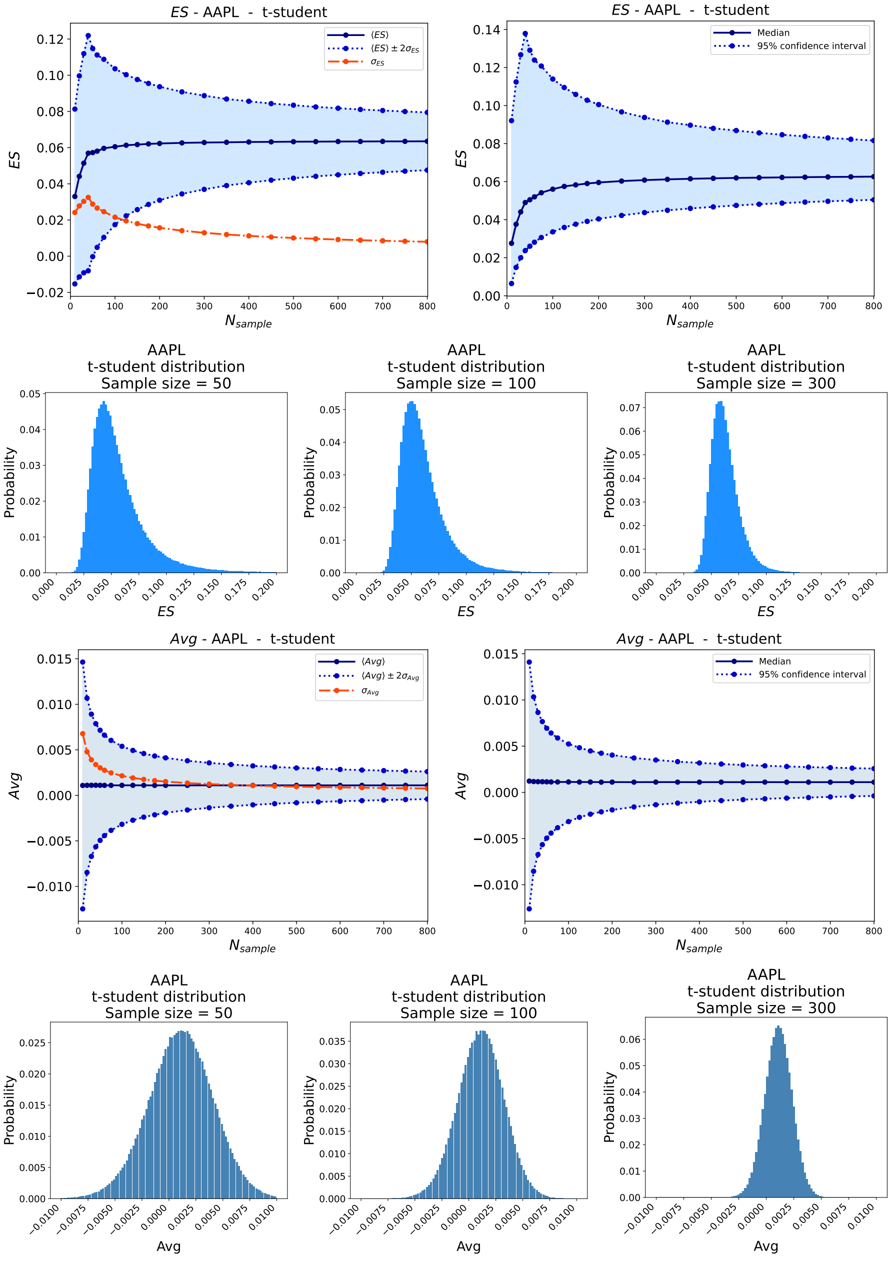

Figure 17: Relationship between the expected shortfall (row 1) and the average (rows 3) of the synthtetic data and the number of data points of the sample size. The left plots present the standard deviation (red) and average plus/minus twice the standard deviation (blue); the right plots present the median and 95% confidence interval. Rows 2 and 4 present corresponding histograms for three different sample sizes.

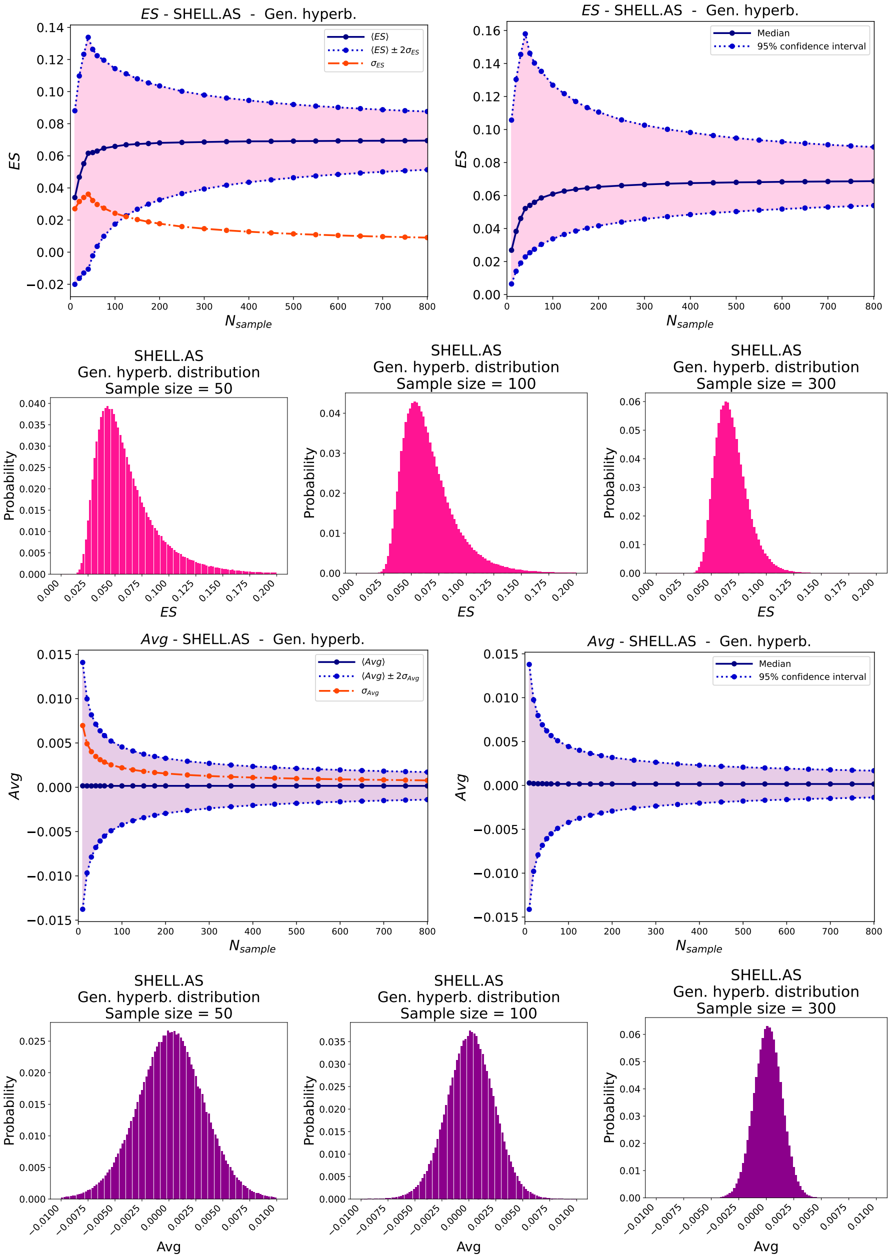

Figure 18: Relationship between the expected shortfall (row 1) and the average (rows 3) of the synthtetic data and the number of data points of the sample size. The left plots present the standard deviation (red) and average plus/minus twice the standard deviation (blue); the right plots present the median and 95% confidence interval. Rows 2 and 4 present corresponding histograms for three different sample sizes.

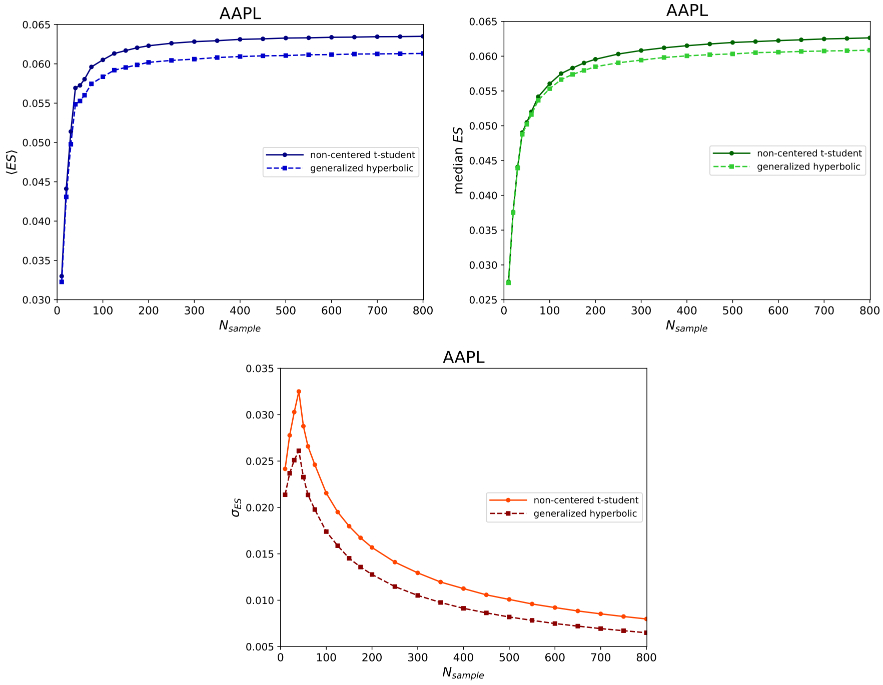

Figure 19: Comparison between the mean, median and standard deviation of the ES for different fitting functions to a given set of observed data (logarithmic daily returns of the AAPL stock).

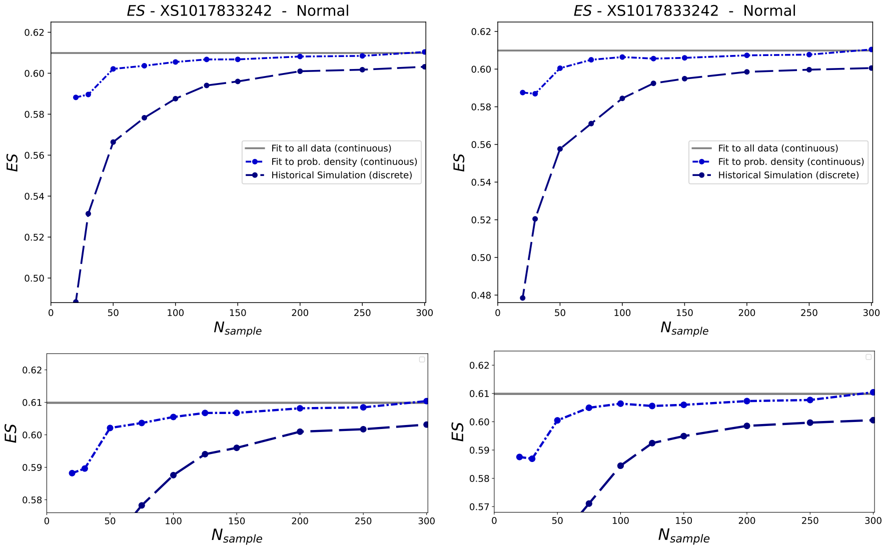

Figure 20: Comparison of the mean (left figures) and median (right figures) of the Expected Shortfall of datasets generated from the distribution parameters obtained from the fitting of a large dataset (absolute returns of a BASF bond, fitted to a normal distribution). The x axis indicates the number of points of the generated datasets (N_{sample}). The dash-dotted lines correspond to historical simulation (discrete); the dashed lines correspond to ES calculated from continuous distributions fitted from each generated dataset to normal distributions. Gray horizontal lines correspond to the ES from the continuous distribution of the large dataset.

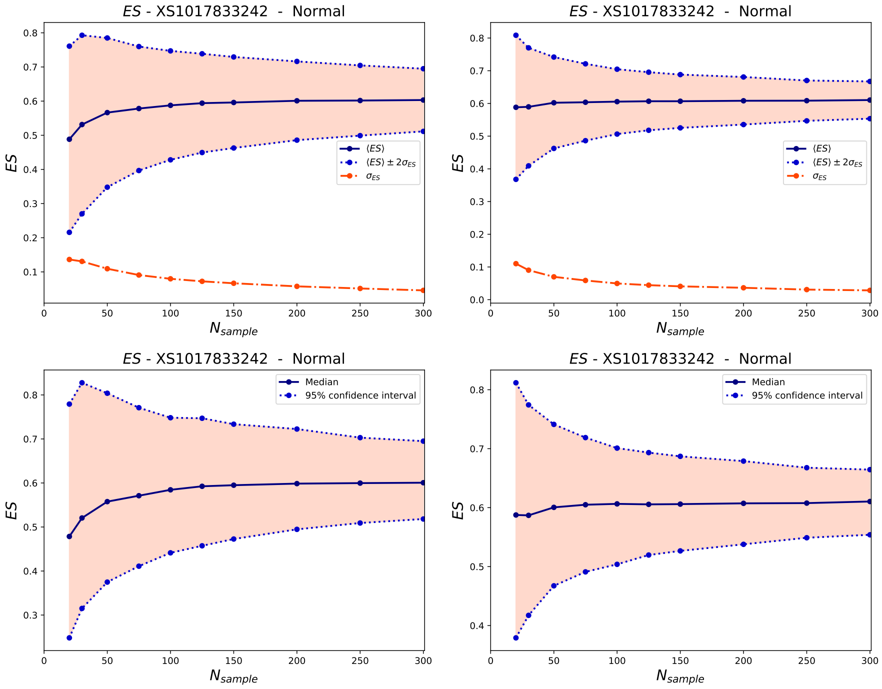

Figure 21: Mean (top) and median (bottom) of the Expected Shortfall from discrete Historical Simulation (left) and from fitting to continuous probability denstiy normal functions of the same datasets, as a function of the size of the dataset. The solid lines correspond to the lines of Fig. 20.

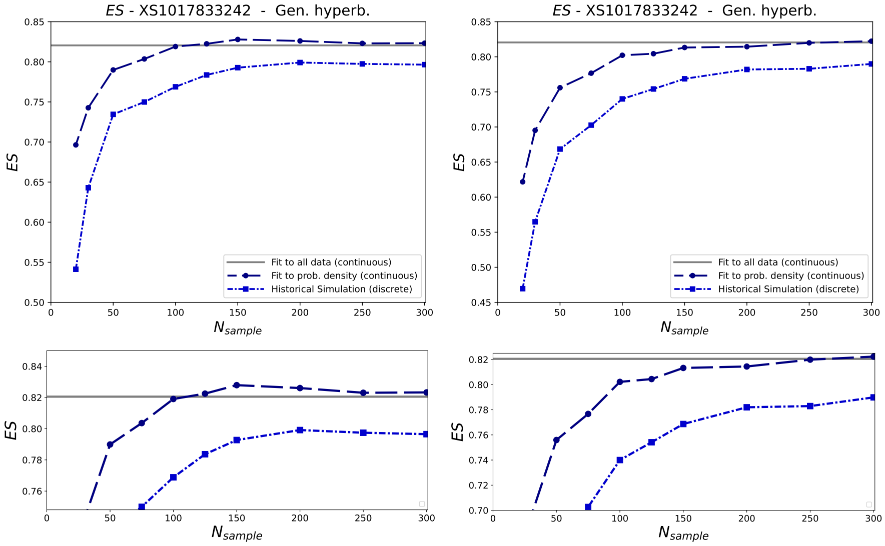

Figure 22: Comparison of the mean (left figures) and median (right figures) of the Expected Shortfall of datasets generated from the distribution parameters obtained from the fitting of a large dataset (absolute returns of a BASF bond, fitted to a generalized hyperbolic distribution). The x axis indicates the number of points of the generated datasets (N_{sample}). The dash-dotted lines correspond to historical simulation (discrete); the dashed lines correspond to ES calculated from continuous distributions fitted from each generated dataset to generalized hyperbolic distributions. Gray horizontal lines correspond to the ES from the continuous distribution of the large dataset.

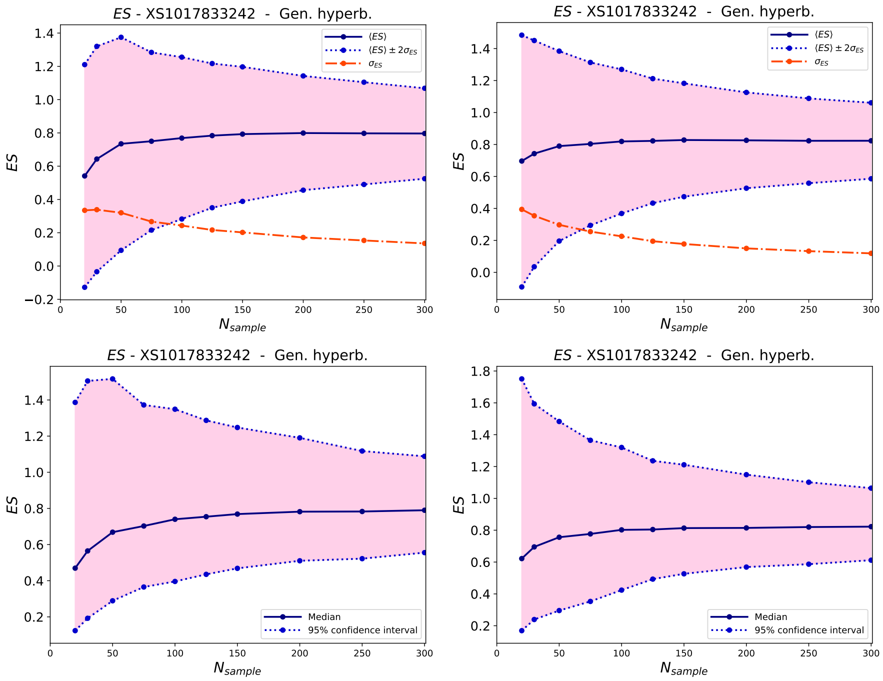

Figure 23: Mean (top) and median (bottom) of the Expected Shortfall from discrete Historical Calculations (left) and from fitting to continuous probability density generalized hyperbolic functions of the same datasets, as a function of the size of the dataset (right). The solid lines correspond to the lines of Fig. 22.

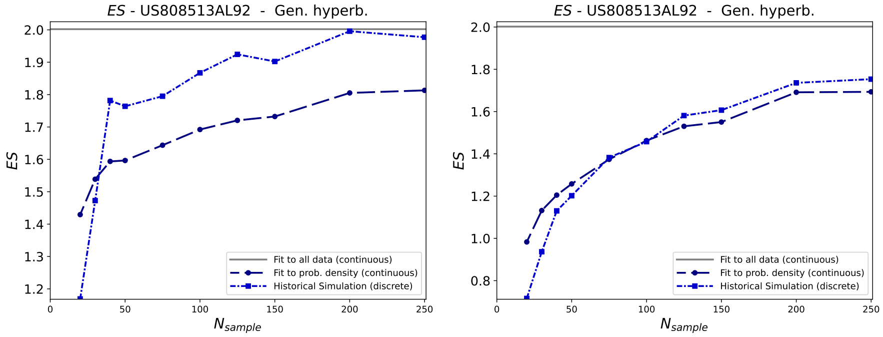

Figure 24: Comparison of the mean (left figures) and median (right figures) of the Expected Shortfall of datasets generated from the distribution parameters obtained from the fitting of a large dataset. The x axis indicates the number of points of the generated datasets (N_{sample}). The dash-dotted lines correspond to historical simulation (discrete); the dashed lines correspond to ES calculated from continuous distributions fitted from each generated dataset to generalized hyperbolic distributions. Gray horizontal lines correspond to the ES from the continuous distribution of the large dataset.

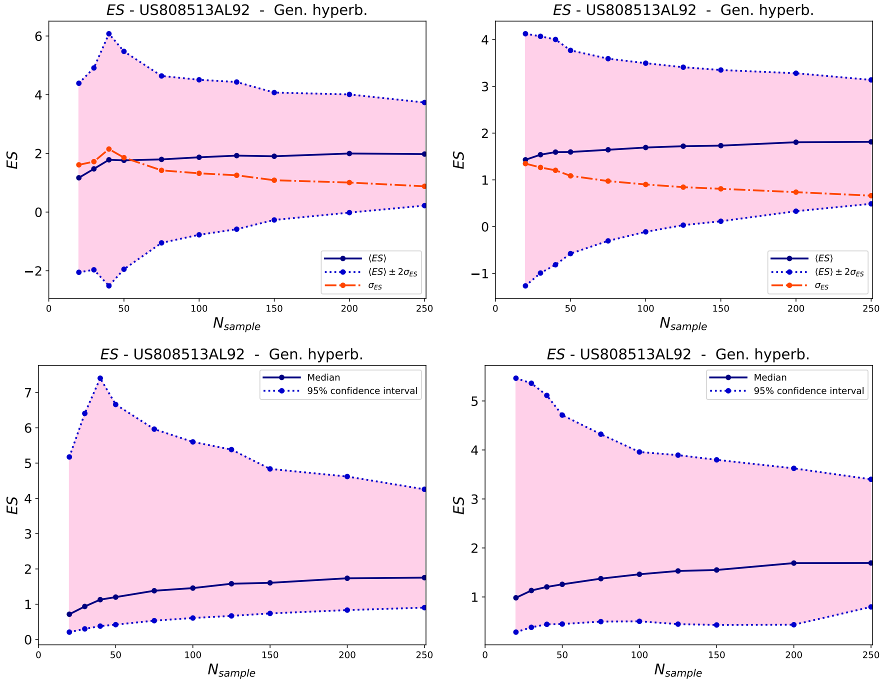

Figure 25: Mean (top) and median (bottom) of the Expected Shortfall from discrete Historical Calculations (left) and from fitting to continuous probability density generalized hyperbolic functions of the same datasets, as a function of the size of the dataset (right). The solid lines correspond to the lines of Fig. 22.

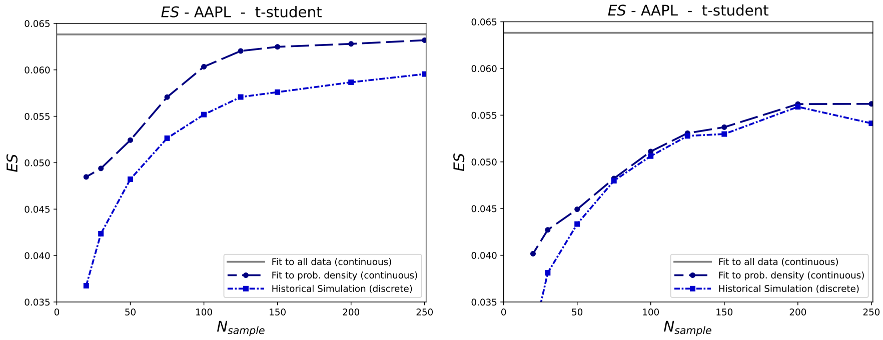

Figure 26: Comparison of the mean (left figures) and median (right figures) of the Expected Shortfall of datasets generated from the distribution parameters obtained from the fitting of a large dataset (apple stock log returns, fitted to a non-centered t-student distribution). The x axis indicates the number of points of the generated datasets (N_{sample}). The dash-dotted lines correspond to historical simulation (discrete); the dashed lines correspond to ES calculated from continuous distributions fitted from each generated dataset to generalized hyperbolic distributions. Gray horizontal lines correspond to the ES from the continuous distribution of the large dataset.

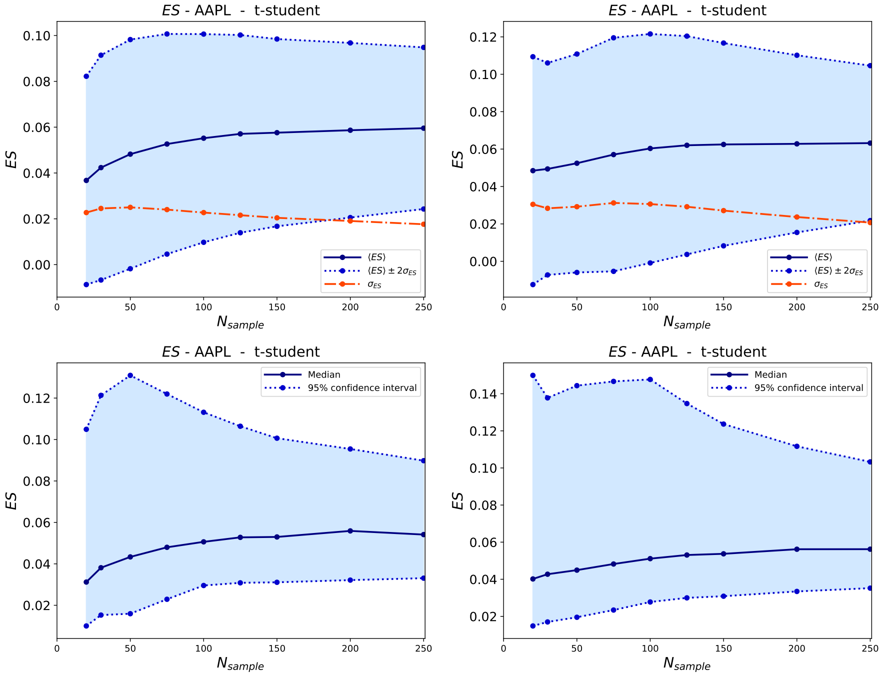

Figure 27: Mean (top) and median (bottom) of the Expected Shortfall from discrete Historical Calculations (left) and from fitting to continuous probability density generalized hyperbolic functions of the same datasets, as a function of the size of the dataset. The solid lines correspond to the lines of Fig. 22.

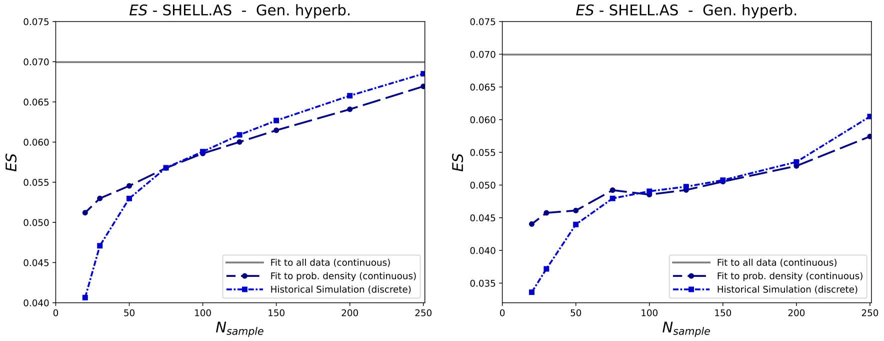

Figure 28: Comparison of the mean (left figures) and median (right figures) of the Expected Shortfall of datasets generated from the distribution parameters obtained from the fitting of a large dataset (log returns of the Shell p.l.c. stock, fitted to generalized hyperbolic distribution). The x axis indicates the number of points of the generated datasets (N_{sample}). The dash-dotted lines correspond to historical simulation (discrete); the dashed lines correspond to ES calculated from continuous distributions fitted from each generated dataset to generalized hyperbolic distributions. Gray horizontal lines correspond to the ES from the continuous distribution of the large dataset.

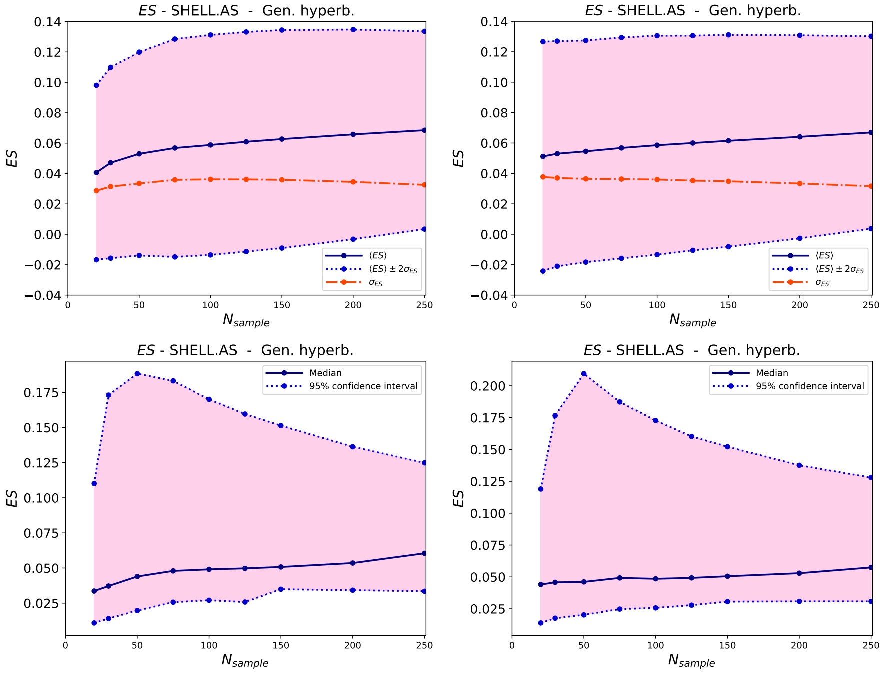

Figure 29: Mean (top) and median (bottom) of the Expected Shortfall from discrete Historical Calculations (left) and from fitting to continuous probability density generalized hyperbolic functions of the same datasets, as a function of the size of the dataset (right). The solid lines correspond to the lines of Fig. 22.