Risk Management Project

Quantitative Asset and Risk Management

Value-at-Risk Analysis of Stock Returns: Historical Simulation, Variance Techniques or Tail Index Estimation?

Master of Science (MSc) in Finance

Habib Joe
Borel Maxime
Bandelier Axel
Joye Yoann

Prof. Eric Jondeau

TA. Alexandre Pauli

University of Lausanne - 29.05.2020

UNIL | Université de Lausanne
HEC Lausanne

Contents

List of Figures

List of Tables

Abstract

Our study is based on the paper " *Value-at-Risk Analysis of Stock Returns Historical Simulation, Variance Techniques or Tail Index Estimation?*" introduced by *Rob van den Goorbergh* and *Peter Vlaar* in 1999. The purpose of this study is to present Value-at-Risk techniques and to evaluate the adequacy of our different models. This paper is based on the data from the Swiss Market Index (SMI) from December 31, 1990 until April 29, 2020 in order to capture the 2008 Global Financial Crisis (GFC) and the COVID-19 crisis. As part of this study, we will present three techniques: (1) Variance techniques, (2) Historical simulation, (3) Tail index estimators. One of the most important conclusions of our study is the benefit of introducing a model with time-varying volatility (GARCH model) and the necessity to implement a distribution compatible with the characteristics of financial data (fat-tails) such as the Student Distribution.

1 Introduction

To begin with, we need to distinguish three different types of risk that can occur in the finance industry : business risk, strategic risk and financial risk. We will focus our analysis on the latter. Pursuing, we can decompose the financial risk into different categories including the market risk which is the one we will work with in this project. Essentially, the market risk comes from the fluctuations in securities' prices. As a measure of that risk, we will use the Value-at-Risk (VaR) measure. This is the maximum loss that can be reached over a certain period of time for a given confidence level. This threshold is represented by the confidence level α. For example, A VaR of 95% is the upper estimate of losses which will exceed the threshold with a probability of 5%. In the risk management field, the VaR is commonly used as a benchmark to measure market risk. However, the risk measure is just a quantile of the distribution, hence, it does not give any information about what happens in the θ% case.

We will start our analysis with the regulatory framework that has been implemented over the last 40 years, since the first Basel Accord. This regulation has set boundaries for the maximum risk financial institutions have been allowed to undertake ever since. Then, we will move on to the statistical evaluation of the data collected. For this, we will use two different tests : Kupiec (unconditional coverage - 1995) and Christoffersen (conditional coverage - 1998). Then, we will go through the data chosen to perform the models. Next, we will explain the different methods used to perform those tests (Variance methods, Historical simulation and Extreme value theory). For each method, we use different assumptions to estimate the VaR. In the discussion, we will cross-compare the different models and methodologies used based on both Kupiec and Christoffersen back-tests.

1.1 Regulation

In July 1988, Basel I, a treaty signed between members of the G10 plus Switzerland and Luxembourg, aimed at fixing the minimum capital requirement that financial institutions need to satisfy. Then, within each institution, there is a different regulatory regime for each and every asset classes. Thus, different reserve of capital required are : 0%, 10%, 20%, 50% and 100% depending on the riskiness of the asset class. Under Basel Accords, systemic banks are required to operate under an asset risk level of 8% of the total capital.

Basel II differs from Basel I in that it introduces credit risk requirements. Since then, these requirements are included in the minimum compulsory capital ratios. The Basel II accord features three pillars : Minimum Capital Requirements, Regulatory Supervision and Market Discipline. It has been established in 2007.

After the Global Financial Crisis and the excessive use of leverage coupled with the lack of liquidity that banks like Lehman Brothers faced in the fall of 2008, the Basel committee had to intervene. In 2009, Basel III was introduced. Essentially, they toughen the requirements. Namely, they fragmented the minimum capital requirements into different financial ratios. These are: Tier 1 capital, which provides a reserve of core capital for operating continuity under liquidity distress (e.g. significant losses), it must not be lower than 10.5%; Tier 2: refers to the additional capital (on top of the core capital) like undisclosed reserve and subordinated of at least 5 years of maturity, it must be at least equal to

2%. The total capital required is estimated by adding up the two tiers. This total capital must be higher or equal to 12.9 % (2% for Tier 2 capital and 10.9% for Tier 1 capital). The other area in which the Basel Accords extended their requirements are related to the systemic importance of the financial institutions. More specifically, banks must set aside a varying amount of capital in accordance with their systemic importance and the economic conditions they operate in. Finally, Basel III, established rules in terms of leverage to prevent excessive borrowings and liquidity measures to ensure that they also have enough cash in hand to face a financial distress, like a bank run. We can cite the leverage (Tier 1 capital / net tangible assets) that must not exceed 3%. Hence, the VaR is a measure that is use to determine the capital requirement according to the Internal Model Approach.

These regulatory restrictions will have an impact in investing portfolio of a financial institution. Indeed, as the risk of the portfolio increases, we have increasing capital requirement too. Capital requirements and VaR measures imposed since Basel II created buffers to protect against market risk.

1.2 Statistical evaluation

In this section, we will present statistical evaluations to judge the adequacy of our models presented in the next section. In order to perform these statistical evaluations, we must first decompose our data sample into two parts:

- *Estimation Sample*: We estimate the parameters of the models in this sub-sample and we predict the Value-at-Risk (VaR) of the portfolio.
- *Evaluation Sample*: A back-test is performed in order to judge the adequacy of the models.

Model estimates are presented in the next section. However, in this chapter, we will present the statistical methodology to assess the adequacy of our models. We are therefore interested in the estimation sample. In order to carry out these statistical evaluations, we will first present and explain the test procedure of *Christoffersen* which is composed of a sequence of two tests:

1. **Test of unconditional coverage (Kupiec test)**
2. **Test of independence**

In other words, this test is called the test for **Conditional Coverage**. In this study, we improve the measurement of adequacy of the models compared to the paper introduced by *Rob van den Goorbergh* and *Peter Vlaar* as we perform a conditional coverage test, which also takes into account the independence test as well.

1.2.1 Test of unconditional coverage - Kupiec (1995)

The Kupiec test is very interesting because it also reflects the back-test proposed by *Basel Accords* for the *Internal Model approach.* This common test of VaR models count the number of time there is VaR exceedances. In other words, it counts the number of days in the evaluation sample when the losses are bigger than the VaR estimated.

If we define x as the number of time the losses exceed the VaR and T the total number of days in our evaluation sample, we can define the *failure rate* as x/T. The intuition behind this test is to compare the realized failure rate to the left tail probability denoted p that was used to determine the VaR (if a confidence level of θ is used, the left tail probability $p = 1 - \theta$). Therefore, we don't reject the test if the failure rate is statistically not different from the left tail probability and we reject the test if it differs substantially. We can formalize this in the following way:

$$H_0 : \frac{x}{T} = p \tag{1}$$

$$H_a : \frac{x}{T} \neq p \tag{2}$$

Furthermore, we know that intuitively, the number of VaR violations follows a *Binomial* distribution. Formally:

$$x \sim B(T, p) \tag{3}$$

We also know that when T increases, we can approximate the Binomial distribution by a *Standard Normal Distribution*:

$$z = \frac{x - pT}{\sqrt{p(1-p)T}} \sim N(0, 1) \tag{4}$$

This test is based on the following likelihood ratio statistic:

$$LR_{unc} = 2\left[log((\frac{x}{T})^x(1 - \frac{x}{T})^{T-x}) - log(p^x(1-p)^{T-x})\right] \tag{5}$$

Under the null that the failure rate does not exceed p:

$$LR \sim \chi^2(1) \tag{6}$$

This Kupiec test is very interesting to evaluate the risk management models of a financial institution because it is a bilateral test. Indeed, it means that the test can be rejected if the failure rate is substantially higher than p (which is a restriction imposed by the regulation) but also it can be rejected when the failure rate is lower than p. The latter case means that the financial institution was too conservative in its model which implies a VaR too high. In this case, a too high VaR can induce higher costs. Indeed, the institution will have too large capital requirements and therefore this will result in a higher opportunity cost.

On the other hand, this test does not take into account the notion of dependence when losses exceed the estimated VaR. Indeed, it is quite possible that losses frequently exceed the estimated VaR in a short period of time (several days in a row as we saw in early March due to the COVID-19 crisis) while not rejecting the Kupiec test. However, even if the failure rate does not exceed substantially the probability p, the financial institution can suffer significant losses in a short period of time. Consequently, it is possible that the latter will fail. Thus, Christoffersen (1998) introduced an **independence test** in addition to the Kupiec test in order to form the **Conditional Coverage** test. This procedure presented by Christoffersen will therefore enable to test whether the estimated models take into account the independence of crashes (the financial institution will be able to better manage its risks).

1.2.2 Test of conditional coverage - Christoffersen (1998)

In this subsection, we are going to present the Christoffersen procedure that will complete the test of unconditional coverage. Indeed, Christoffersen implements a second test which is a **test of independence**. In order to perform this test, we have to describe a variable Hit_t as follows:

$$Hit_{t+1} = \begin{Bmatrix} 1 & if & r_{p,t+1} < -Va\bar{R}_{\theta,t} & & Violation \\ 0 & if & r_{p,t+1} > -Va\bar{R}_{\theta,t} & No & Violation \end{Bmatrix} \tag{7}$$

Then, we know that under the alternative the variable Hit_t is a two-state Markov chain, with transition probability matrix Π_1:

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix} \tag{8}$$

With:

- π_{01} corresponding to the probability that $Hit_t = 1$ knowing that $Hit_{t-1} = 0$ (the model is violated in t, however it is not violated in the previous period $t-1$).
- π_{11} corresponding to the probability that $Hit_t = 1$ knowing that $Hit_t = 0$ (the model is violated in t and in the previous period $t-1$ as well).

We estimate the probability π_{ij} by:

$$\hat{\pi}_{ij} = x_{ij}/(x_{i0} + x_{i1}) \tag{9}$$

With x_{ij} corresponding to the number of observations with value i followed by value j. The likelihood function under the alternative is:

$$L(\Pi_1|Hit_t, t = 1, ..., T) = (1-\pi_{01})^{x_{00}}(\pi_{01})^{x_{01}}(1-\pi_{11})^{x_{10}}(\pi_{11})^{x_{11}} \tag{10}$$

We also have to compute the transition probability under the null:

$$\Pi_2 = \begin{bmatrix} 1-\pi_2 & \pi_2 \\ 1-\pi_2 & \pi_2 \end{bmatrix} \tag{11}$$

The likelihood function is the following:

$$L(\Pi_2|Hit_t, t = 1, ..., T) = (1 - \pi_2)^{x_{00}+x_{10}}(\pi_2)^{x_{01}+x_{11}} \tag{12}$$

Now, the probability π_2 is estimated as:

$$\hat{\pi}_2 = (x_{01} + x_{11})/T \tag{13}$$

Therefore, we can compute the likelihood ratio statistic of the independence test as:

$$LR_{CCI} = -2log\left[\frac{L(\hat{\Pi}_2|Hit_t, t = 1, ..., T)}{L(\hat{\Pi}_1|Hit_t, t = 1, ..., T)}\right] \tag{14}$$

As with the Kupiec test, under the null, the likelihood ratio statistic is distributed as $\chi^2(1)$.

Finally, according to the *Christoffersen* procedure, we can compute the **conditional coverage test** which is the sum of the unconditional coverage test (Kupiec test) and the independence test:

$$LR_{cc} = LR_{unc} + LR_{CCI} \tag{15}$$

This test is distributed under the null as a $\chi^2(2)$.

Now, we can see that when the loss exceeds many times the VaR estimated in a short period of time, the likelihood ratio statistic of the independence test increases, and hence, the likelihood ratio for the conditional coverage test increases as well. This test proposed by Christoffersen becomes extremely important in the case where the Kupiec test alone is not rejected but the loss exceeds the VaR many times in a short period of time, leading to a rejection of the conditional coverage test (which wouldn't be rejected if we only had used the unconditional coverage test).

This procedure will be extremely important in order to perform a *backtest* of the models estimated in the following sections.

In terms of implementation, we process as follows :

First, considering the loss function for the evaluation window, we replace all the values bigger than the VaR by 1 and 0 elsewhere.Then we introduce different measures that are : x_{00}, the number of times the model yields "success" then "success" back-to-back; x_{01}, the number of times we have "success" then "fail", x_{10} the number of time when the model results in "fail" then "success", and x_{11} number of time where we have two fails in a row. Then, we compute π, the probability of having a failure at t, π_0, the probability of having a failure at t given a success at t-1, and π_1, the probability to have a failure at t given a failure at t-1 with :

$$\pi = \frac{(x_{01} + x_{11})}{(x_{00} + x_{01} + x_{10} + x_{11})}$$
$$\pi_0 = \frac{x_{01}}{(x_{00} + x_{01})} \tag{16}$$
$$\pi_1 = \frac{x_{11}}{(x_{11} + x_{10})}$$

Finally, we compute the LR_{CCI} using the last parameters as follow :

$$LR_{CCI} = -2\log\left(\frac{(1-\pi)^{x_{00}+x_{10}}\pi^{x_{01}+x_{11}}}{(1-\pi_0)^{x_{00}}\pi_0^{x_{01}}(1-\pi_1)^{x_{10}}\pi_1^{x_{11}}}\right) \tag{17}$$

Please notice that this computation requires a vector of violation with at least one violation of the VaR. In the case that the dummy vector is only 0, the function return NaN. We face this issue for the confidence level for the VaR at 0.01%. As the quantile is far in the tail, it is likely that the VaR for this given level has no violation, hence we are not able to compute the independence test.

2 Data

The data used in this study consists in one index. Value-at-Risk can be computed at the level of the portfolio directly or considering multivariate approach. As the goal of this study is to evaluate the different means of modelling the risk measure, we decide to focus on a unique index. However, we have to keep in mind that, by doing so, we might miss some relation between assets that can be captured by the multivariate approach. We focus on our domestic market and choose to work with the Swiss Market Index (SMI). We expect that the results will differ from the reference paper as we do not use the same data (they used the AEX and DJIA indexes). Moreover, the results should differ also because we are not considering the same time periods than in the paper and neither the same length of sample to estimate our parameters and back test the models. Moreover, compared to the paper, we add some models including the Age-weighted Historical Simulation. We do not use the mixture of Gaussian but the Skewed-t for the dynamic part. We also implement different methodologies for the EVT part. Nevertheless, we expect more or less to find similar results in terms of models.

We use daily data from December 31, 1990 until April 29, 2020, provided by Yahoo Finance. In this particular time interval, it is easier for us to get the data time series from public database. We are aware that the database is not the same quality as from Datastream. Moreover, our time series might suffer from the survivorship bias as we do not take into account the change of components in the index. All closed-market days are removed. We end up with 7381 observations in the sample of price, hence 7381 is the number of observed returns in as shown in Table 1.

In risk management point of view, it was interesting to use the most recent data available. Indeed, the COVID-19 crisis brings on a lot of nervousness on the markets. As we are using model such as GARCH for modelling the volatility, we will be able to assess its capability to deal with with very unstable conditions.

As in the paper, we split the sample in two subsamples, namely the estimation sample and the evaluation sample. The estimation sample start in January 1, 1990 to December 31, 2006. We use this sample to estimate the parameters. This methodology assumes that the parameters are not impacted by events in the evaluation window, which is not really realistic. Regarding the evaluation window, it starts from January 1, 2007 to April 29, 2020. We choose a back-test period that includes the 2007-09 Global Financial Crisis. In our sense, the length of the former is relatively long compare to the latter. Hence, we think that it might create some problems regarding the back-testing importance. Unfortunately, we are not able to have longer sample from our data source.

Finally, we use the log returns in all our computations. This generates a more conservative approach than using simple returns as they are always smaller. In the Table 1 and Table 2 below, we inspect the characteristics of the time-series by computing descriptive statistics and auto-correlation on the whole sample.

In this paragraph, we refer to Table 1. Regarding the measures of asymmetry and tailedness, we clearly observe a non-normal distribution for the return (we have to statistically test it with a Jarque Bera test or a Lilliefors test). Hence, we expect that model using normal distribution to give bad results in term of VaR back-testing as the distribution is negatively skewed and has excess kurtosis. The model using Hansen Skewed-t or Extreme Value Theory should give better results.

Observations	Values
Number of observations	7380
Mean	0.0003
Median	0.0007
Max	0.1079
Min	-0.1013
Standard deviation	0.0114
Skewness	-0.3267
Excess Kurtosis	7.2826

Table 1: Descriptive Statistics - SMI

Lags	ACF	PACF	Ljung-Box	pValue
1	0.3029	0.3029	677.5617	0
2	0.2768	0.2037	1243.4206	0
3	0.2700	0.1625	1781.8669	0
4	0.2537	0.1212	2257.3270	0
5	0.2589	0.1165	2752.4838	0
6	0.1785	0.0066	2987.9276	0
7	0.2019	0.0569	3289.1898	0
8	0.2010	0.0561	3587.8689	0
9	0.1899	0.0444	3854.3641	0
10	0.1991	0.0575	4147.3405	0

Table 2: Auto-correlation in ε_t^2 - SMI

In term of auto-correlation, according to Table 2, we obtain standard results for the ε_t^2. As expected, we have some predictability in squared unexpected returns. We always reject the null hypothesis of no auto-correlation. Doing a Ljungbox test on squared unexpected returns is an alternative way to test for ARCH effect instead of doing a LM test. We definitely need a model for the volatility. Hence we expect that models with constant mean will perform poorly compare to dynamic ones.

Finally, in Figure 1, we see all the returns overtime. We clearly see graphically the volatility clustering effect is suggesting even more the need for a model such as a GARCH.

3 Variance methods

In this sub-section, we will present the variance methods for estimating VaR when making assumptions about the distribution of returns in our data. The variance of returns is directly related to the VaR. Indeed, the higher the variance, the more likely it is to have an equally high VaR. In this chapter, we will look at two different models:

- **Static Model**: This naive model ignores the volatility clustering.
- **Dynamic Model**: This more advanced model takes into account the volatility clustering (GARCH model).

For each of these models we will assume the following distributions ranging from the least complex to the most flexible and realistic:

- **Normal Distribution**: Naive assumption that does not take into account asymmetry and fat-tails in returns.
- **Student-t Distribution**: More complete assumption that does take into account fat-tails in returns. However, it ignores the negative asymmetry in returns which is generally the case when we talk about financial returns.
- **Skewed Generalized-t Distribution**: More advanced model that captures fat-tails as well as the asymmetry in returns.

3.1 Static Models

In this subsection, we are going to present static models that does not capture the temporal variability of the volatility. These are very simple models that does not adequately represent reality. Indeed, these models are unable to adapt to new market conditions (particularly when volatility rises sharply) and can therefore quickly become obsolete. This can be extremely problematic if we take the regulator's point of view. Indeed, it would be possible that in an event of high market volatility, the financial institution may not have provided the necessary capital requirements to cope the financial distress.

Throughout this work, we're going to use several levels of left tail probability p from the most conservative level of 0.001% to a more reasonable and flexible level of 5%. However, the most conservative level of 0.001% is not very interesting because events involving this level are extremely rare. We will also carry out a back-test proposed by Christoffersen (1998). This method was presented in the introduction.

3.1.1 Normality

Within the framework of this first hypothesis, we will assume that the returns composing our data follow a Normal distribution. This hypothesis is very simplistic and unrealistic because it does not take into account the asymmetry in the distribution (negative Skewness) and the fat-tails (positive excess Kurtosis) as shown in Table 2. However, the Normal distribution has important properties that helps to facilitate the computation of the VaR estimate. First, the parameters of this distribution are easy to estimate as it exists a close form analytical solution. Second, the normal distribution has the characteristic of additivity which means that the sum of two Normal distributions is also normal. This characteristic is essential to compute multi-day VaRs based on one-day VaRs. In fact, we have the following expression which is called the *square root of time rule*:

$$VaR^{(T)} \approx \sqrt{T} VaR^{(1)} \tag{18}$$

We estimate the parameter of a normal distribution using the *maximum likelihood method* in Table 3.

Parameter	Estimate	t-stat
μ	0.0005	2.5755
σ	0.0113	89.8387

Table 3: Static Normal - Parameters estimated

As we assume that the returns are independent and identically normally distributed, we can compute the Value-at-Risk in a similar fashion as proposed in the original paper:

$$VaR = -W_0(\exp(\mu + \sigma\Phi^{-1}(p)) - 1) \tag{19}$$

With:

- W_0: Initial value of the portfolio.
- $\Phi()$: Cumulative distribution function of a $N(0,1)$.

p	VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.7961	162	0.1686	39.7133†	39.8820†	5.991
1%	2.5493	71†	32.2462	42.2372†	74.4834†	5.991
0.5%	2.8236	57†	59.7687	48.9909†	108.7596†	5.991
0.1%	3.3867	37†	110.9259	22.8423†	133.7683†	5.991
0.01%	4.0706	22†	141.0267	20.2903†	161.3171†	5.991

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence is distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance of 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 4: Static Normal - VaR estimate back-test

In Table 4, we estimate the VaRs for different left tail probabilities p and perform a back-test using Christoffersen procedure. We can notice that our model performs poorly. Indeed, we observe that we reject the conditional test (Christoffersen test) for all probabilities. Concerning the Kupiec test, we also reject all the left tail probabilities except the higher probability : $p = 5\%$. We can conclude that the static normal model is not a good one.

In Figure 2, we plot the daily returns from January 2008 to January 2020 with the static VaR at 5%. We can clearly see that the loss exceeds many times the threshold (which is the VaR estimated at 5%). We can also notice that the VaR is always constant. In fact, this naive static model implies that the VaR does not adjust when the volatility increases which results in a very inefficient model.

3.1.2 The Student-t Distribution

In this subsection, we make the assumption that the returns are following a Student-t Distribution. This assumption is more realistic than the Normality assumption simply because the Student-t distribution can adjust for the fat-tails of our returns. Indeed, this distribution takes into account that returns have a positive excess Kurtosis. However, it does not take into account the asymmetry in the empirical distribution (negative Skewness). In Table 5, we estimate the parameters of the distribution using the Maximum Likelihood method. We can notice that the ν parameter representing the degree of freedom of the distribution is very low. This is an indicator that the distribution has fat-tails (which is not the case under the normality assumption). In the limit case as ν goes to infinity, the student-t converge to a Normal distribution. Hence, with this distribution, we expect to obtain larger VaR estimates than with the normal as the quantile of the student-t must be bigger in absolute value.

We compute the VaR as follows:

$$VaR = -W_0(\exp(\mu + \sigma F_\nu^{-1}(p)) - 1) \tag{20}$$

With:

- $F_\nu()$: Cumulative distribution function of a standardized t-distribution.

Parameter	Estimate	t-stat
μ	0.0007	5.0092
σ	0.0078	49.7965
ν	3.6218	16.4111

Table 5: Static Student-t - Parameters estimated

p	VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.6195	200†	Inf	Inf†	Inf†	5.991
1%	2.9663	49†	6.4053	50.7889†	57.1942†	5.991
0.5%	3.6994	28†	6.3586	16.2340†	22.5927†	5.991
0.1%	5.9439	5	0.7125	8.2299†	8.9424†	5.991
0.01%	11.1585	0	0.6686	NaN	NaN	5.991

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence is distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance at 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 6: Static Student-t - VaR estimate back-test

In Table 6, we can see that we reject the conditional coverage test for all left tail probabilities. However, we do not reject the lower probabilities ($p = 0.1\%, 0.01\%$) for the Kupiec test. It is better than the static normal model but it is not sufficient to be a good one. As mention above, the VaR estimates are larger than with the Normal except at 5%. In our sense, it might come from a much lower standard deviation for the student-t than the Normal.

In Figure 2, as before, we plot the daily returns from January 2008 to January 2020 with the static VaR at 5%.

3.1.3 Skewed Generalized-t Distribution

In this subsection, we now assume that the returns are following a Skewed Generalized-t Distribution which is a more complex and realistic assumption. Indeed, this distribution allow to take into account two important characteristics of our returns:

- Negative Skewness
- Positive excess Kurtosis

In Table 7, we estimate the parameters of the distribution using the Maximum likelihood method. To perform this, we have to implement our own likelihood function and to use an optimizer as Matlab does not directly allow to estimate the Skewed-t. We use a standard methodology (the function is available is the document "LLskewt").

Parameter	Estimate	t-stat
μ	0.0005	0.6154
σ	0.0001	32.9979
ν	3.6392	699.0063
λ	-0.0613	-2.7312

Table 7: Static Skewed-t - Parameters estimated

Parameters are satisfying as they capture the asymmetry and the fat-tailedness at significant level. We notice a very low σ which will impact the the VaR computation.
We compute the VaR as follows:

$$VaR = -W_0(\exp{(\mu + \sigma G^{-1}_{\nu,\lambda}(p))} - 1) \tag{21}$$

With:

- $G_{\nu,\lambda}()$: Cumulative distribution function of a Skewed-t distribution.

p	VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.6969	180	1.0155	50.6594†	51.6749†	5.991
1%	3.1348	42	2.0523	31.4382†	33.4906†	5.991
0.5%	3.9174	24	2.8098	18.8067†	21.6165†	5.991
0.1%	6.3091	4	0.1215	0.0096	0.1311	5.991
0.01%	11.8388	0	0.6686	NaN	NaN	5.991

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence is distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance at 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 8: Static Skewed-t - VaR estimate back-test

In Table 8, we can notice that we reject the Christoffersen test for every left tail probabilities except for $p = 0.1\%$. Nevertheless, the unconditional test is not rejected for every probabilities. It shows us that, compared to the other statics models, the skewed-t is the best.

In Figure 2, we can plot the daily returns from January 2008 and January 2020 with the static VaR at 5%.

3.2 Dynamic Models (GARCH)

In this subsection, we will present a much more comprehensive and realistic model. Indeed, we have seen that in the static models, we do not take into account the phenomenon of the volatility clustering. Therefore, we are going to present the GARCH (*Generalised Autoregressive Conditional Heteroskedasticity*) model presented by *Bollerslev* (1986). In this model, we first assume the innovation process ϵ_t as:

$$\epsilon_t = \sigma_t z_t \tag{22}$$

With $z_t \sim iid(0, 1)$. In the GARCH model, we assume that the conditional variance of the innovations $\sigma_t^2 = E(\epsilon_t^2 | I_{t-1})$ depends on previous squared innovations. The information set I_t includes all the innovations at time t and before. The GARCH model helps to capture the serial correlation of volatility. A $GARCH(p, q)$ model can be written as following:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{23}$$

As we can see, p lags are included in the squared innovations and q lags in the conditional variance. In this model, we assume the returns to follow the process:

$$r_t = \mu + \epsilon_t \tag{24}$$

With μ corresponding to the unconditional mean of the process. In the course of our work, we're going to assume a $GARCH(1, 1)$ (which is sufficient to capture the serial correlation of volatility).The model captures the intensity of the shocks and the persistence of the volatility. Hence:

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{25}$$

We need to add some constraints on the parameters: $\omega > 0, \alpha \geq 0, \beta \geq 0$ in order to be sure to have a strictly positive variance. We can interpret the $GARCH(1, 1)$ model as follows: if the previous day the volatility was high, we will have a high volatility today (can be amplifi or reduce depending on the innovation process) and vice versa. This phenomenon depends on the parameters $\alpha and \beta$. We can also prove that the unconditional variance can be written as:

$$\sigma^2 = \frac{\omega}{(1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j)} \tag{26}$$

Therefore, we can notice the following property: if $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ and ω is a finite number, then the process ϵ_t is covariance stationary. If we assume a $GARCH(1, 1)$ model, we need to have $\gamma = \alpha + \beta < 1$ in order to have a covariance stationary process. Please note that γ capture the persistence of our model:

- if $\gamma < 1$: conditional variance has the property of a mean-reversion parameter.

- if $\gamma = 1$: We have persistence in our model. Hence, the unconditional variance explodes and goes to infinity. Nevertheless, the process is still strictly stationary.

Finally, in order to estimate GARCH parameters, we use the QMLE approach with the normal distribution. This allows us to have consistent parameters but not efficient. However, using the Normal avoids having inconsistency because of misspecifying the distribution.

3.2.1 Normality

In this subsection, we first assume that the distribution of the index return innovation ϵ_{t+1} follows a Gaussian distribution. In order to implement this, we assume that $z_t \sim iidN(0,1)$ so that:

$$\epsilon_{t+1}|I_t \sim N(0, \sigma_t^2) \tag{27}$$

We then estimate the parameters by maximum likelihood. The results are summarised in Table 9:

Parameter	Estimate	t-stat
μ	0.0001	2.5755
ω	0.0000	2.0222
α	0.1251	5.2384
β	0.8314	21.3062
$\alpha + \beta$	0.9564	-2.1889

Please notice that we estimate the μ parameter as the unconditional mean (simple average).

Table 9: Dynamic Normal - Parameters estimated

We can see in the above table that the parameters are satisfying. We notice that the persistence of the volatility is significant but is the lowest persistence of all GARCH processes.

Secondly, as in the static model, we perform a back-test. However, as the estimation of the VaR changes through time, we summarize the average VaR in Table 10:

As we can notice in the previous table, this dynamic model assuming normality performs poorly in a risk management point of view. First of all, we can see that the unconditional coverage test and conditional coverage test are always rejected for every left tail probabilities. In conclusion, we can say that this model does not fit our empirical data at all. Indeed, the normality assumption does not take into account the asymmetry and the fat-tails of the innovation process even if the GARCH generates fat-tails in the unconditional distribution. The results are a bit less worse than with the model without the GARCH except at 5%.

In Figure 3, we can plot the daily returns and daily VaRs (level of 5%) from July 2008 to April 2020. First of all, we can notice that the dynamic VaR (induced by GARCH model) allows to adjust in case of high market volatility. Indeed, we can see that in March 2020 (during COVID-19 outbreak), the VaR increases a lot to adjust due to the spike in the volatility. However, as we can see in this

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.6293	204†	Inf	NaN	NaN	5.991
1%	2.3134	68†	27.7897	6.0187†	33.8084†	5.991
0.5%	2.5623	51†	45.5676	7.0780†	52.6455†	5.991
0.1%	3.0733	25†	57.4271	6.1255†	63.5526†	5.991
0.01%	3.6934	13†	69.8940	4.2130†	74.1070†	5.991

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence is distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance at 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 10: Dynamic Normal - VaR estimate back-test

figure, the normality assumption is not realistic which leads to a very poor model in terms of risk management point of view.

3.2.2 The Student-t Distribution

As part of this subsection, we make a more realistic assumption about the distribution of innovations. In fact, we assume that their process follows a Student-t distribution. This assumption makes it possible to take into account the fat-tails of the return distribution. As a result, we believe that this dynamic model can perform better than when we assumed the normality assumption. In Table 11, we estimate the parameters using the Maximum Likelihood method:

Parameter	Estimate	t-stat
μ	0.0001	2.5755
ω	0.0000	3.3163
α	0.1003	5.9183
β	0.8813	43.2583
$\alpha + \beta$	0.9816	-2.7342
ν	8.4252	6.1998

Please notice that we estimate the μ parameter as the unconditional mean (simple average).

Table 11: Dynamic Student-t - Parameters estimated

We observe that the process is a slightly more persistent as suggest by the sum of α and β. We also remark that the degree of freedom increases suggesting the fat-tails are less pronounced in the innovation process.

As in the previous cases, we perform a back-test in Table 12. As we can notice, our model perform very well when we take into account the fat-tails characteristics of the financial returns. Indeed, we can see that we do not reject the Kupiec test for any left tail probabilities p. However, we still reject the independence test for $p = 5\%, 1\%, 0.1\%$ and thus we also reject the conditional coverage test. It suggests that the violation occurs to close to each others. This dynamic model with Student-t innovations perform better than the dynamic model with normal innovations and the static model with Student-t assumption.

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.8207	158	0.5366	16.5324†	17.0690†	5.991
1%	2.8272	38	0.6044	11.3719†	11.9762†	5.991
0.5%	3.2644	19	0.3006	2.7528	3.0534	5.991
0.1%	4.3347	4	0.1215	9.2364†	9.3579†	5.991
0.01%	6.0875	1	0.8602	0.0006	0.8608	5.991

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence is distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance at 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 12: Dynamic Student-t - VaR estimate back-test

In Figure 3, we plot the dynamic of VaRs estimated for a Dynamic Student-t (with $p = 5\%$). We compare this evolution with the dynamic of VaRs for a Dynamic Normal. As we can see, in a risk management point of view, the assumption of Student-t innovations is way more efficient than the Normal one. However, the increase in VaRs is more expensive for the financial institution using this method.

3.2.3 Skewed Generalized-t Distribution

In this subsection, we present the most advanced and sophisticated model as part of the variance techniques. In fact, we now assume that the innovation process follows a Skewed Generalized-t distribution. This distribution allows to take into account the fatness in the tails and the asymmetry in the innovation distribution (Skewness) which is the case in our data sample. We therefore believe that this dynamic model will be the best among the others presented earlier in this section. In Table 13, we estimate the different parameters of this dynamic model:

Parameter	Estimate	t-stat
μ	0.0001	2.5755
ω	0.0000	3.9377
α	0.1033	6.6123
β	0.8787	50.0669
$\alpha + \beta$	0.9820	-3.0455
ν	8.7462	6.2481
λ	-0.1305	-5.7557

Please notice that we estimate the μ parameter as the unconditional mean (simple average).

Table 13: Dynamic Skewed-t - Parameters estimated

The estimated parameters are close to the student-t in terms of persistence and ν. We see that we have a negative skewness in the innovation process. All the parameters are significant.

In Table 14, we perform a back-test for this specific model. Quite surprisingly, this dynamic model does not perform very well. In fact, we do not reject the Kupiec test only for $p = 0.01\%$. Similarly, we do not reject the independence test for $p = 0.01\%$, hence, we also do not reject the test of conditional coverage for this left tail probability. As a result, this model is not better than the dynamic Student-t model which performs way better. This can be explained by the fact that the Skewed-t takes into account the negative skewness of the financial data at the expense of a bigger excess kurtosis. Therefore, it is more important to take correctly into account the fat-tails of the distribution instead of the negative asymmetry characteristic of the returns. We illustrate that graphically in Figure 13. The difference might come from the fact that we do not use the appropriate distribution for the Swiss index. In order to verify that, one can perform a Diebold, Gunther and Tay (DGT) test. If the result shows that the skewed-t is not in adequacy, it explains the results.

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.6680	199†	Inf	Inf†	Inf†	5.991
1%	2.6487	48†	5.6520	7.9225†	13.5744†	5.991
0.5%	3.0744	27†	5.3565	5.5430†	10.8994†	5.991
0.1%	4.1130	9†	6.5220	5.7088†	12.2308†	5.991
0.01%	5.8013	1	0.8602	0.0006	0.8608	5.991

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence is distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance at 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 14: Dynamic Skewed-t - VaR estimate back-test

In Figure 3, we plot the evolution of the estimated VaRs from July 2008 to April 2020 for a level $p = 5\%$. As we can notice, our dynamic model adapts very well to the volatility changing through time. We can see that the losses do not exceed many times the VaRs estimated (indeed, the Kupiec test is not rejected). However, we can see that during the COVID-19 outbreak, the losses exceed many times the VaRs in a very short period of time which leads to a rejection of the independence test.

4 Historical simulation

The Historical simulation (or HS) is another approach to evaluate the VaR. The advantages of this method is that it does not have any distributional prerequisites and it doesn't include any parameter to estimate. In fact, the HS uses a loss empirical cumulative distribution of data (portfolio returns). The procedure of this method is the following. The sample are grouped in equally long *windows* (subsamples). The number of portfolio is $T-n+1$. T is the sample size and n is the rolling window size. To estimate the VaR, we pick an identical quantile for each subsample (pth), R_t^p. In order to compute the VaR. With this procedure, we are able to take into account the change in the true empirical cumulative distribution as we use a rolling window. We use the following formula:

$$\widehat{VaR}_{t+1|t} = -W_0 R_t^p \tag{28}$$

We take the portfolio returns at time t and the $n-1$ preceding returns to obtain the VaR_{t+1}.

In the following two subsections, we have to keep in mind that the estimation of the VaR_{t+1} depends on the size of the rolling window. In particular, we cannot compute the risk measure for confidence levels higher than 1/(size of rolling window). Meaning that, if we have a rolling window of 500 past returns or less, we are not able to provide a good estimation for confidence level 0.1%. Moreover, we do not have enough returns in our rolling windows to compute a probability of 0.01% for each model presented in this section.

The critical assumption with the Historical Simulation is the size of the rolling window. With small rolling window, the risk measure is very sensitive to large outcomes. A longer window keep old returns in memory which can be irrelevant in the current market situation.

4.1 Standard

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.7150	195†	Inf	35.3569†	Inf†	5.991
1%	2.8881	53†	9.8248	10.1999†	20.0247†	5.991
0.5%	3.6303	31†	9.7876	9.1410†	18.9286†	5.991
0.1%	-	-	-	-	-	-
0.01%	-	-	-	-	-	-

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence are distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance of 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rates, we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 15: HS - N=250 - VaR estimate back-test

In Table 15, for a Historical Simulation with a subsample of 250 observations, we can observe that the unconditional test, Kupiec, rejects the null hypothesis that the failure rate is equal to the VaR percentile for all level of p. Similarly, concerning the Christoffersen test, we clearly see that, for every levels of p, we reject the test. We can conclude that this very simple and naive technique perform poorly.

Regarding the Figure 5, we can see that with a lower rolling window, we match better the evolution of the VaR over time than with longer rolling window. As the length of the rolling window increases it becomes more and more like the static model. With this methodology, every returns in the rolling window has the same weight, 1/n.

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.6980	197†	Inf	Inf†	Inf†	5.991
1%	3.1858	49†	6.4053	26.3311†	32.7364†	5.991
0.5%	3.6059	34†	13.8036	13.1030†	26.9066†	5.991
0.1%	-	-	-	-	-	-
0.01%	-	-	-	-	-	-

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence are distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance of 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 16: HS - N=500 - VaR estimate back-test

In Table 16, as in the previous table , we have the same results: a rejection for the unconditional test (Kupiec) for every left tail probabilities p and a rejection for the conditional test (Christoffersen) for every left tail probabilities. It implies that, increasing the number of observations from 250 to 500, is not enough to improve the model.

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.7270	194†	4.2405	50.5502†	54.7908†	5.991
1%	3.1874	48†	5.6520	32.7522†	38.4041†	5.991
0.5%	4.0272	27†	5.3565	38.7781†	44.1346†	5.991
0.1%	6.1227	11†	10.9062	4.8844†	15.7906†	5.991
0.01%	-	-	-	-	-	-

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence are distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance of 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 17: HS - N=1000 - VaR estimate back-test

Yet, increasing the number of observations to 1000 per window does not change the results obtained in the first two cases. Indeed, we reject for every p for the unconditional test. We still have a rejection for conditional test for every left tail probabilities.

We can see in the Figure 4 that the larger the number of observation is, the more stable is the 5% VaR's quantile. This is intuitive since the number of observation after the VaR is much larger with a subsample of 1000 observations than with 250. Therefore, the rolling window is larger and, as a result, fewer observation are falling outside the tail after the VaR. That is, smaller windows (like 250 observations) tend to eject quickly a larger number of return points than with a window of 1000 observations. In other words, the duration of the effect of extreme values on the VaR is greater with small windows than with larger ones.

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.8893	150	1.9156	37.2781†	39.1936†	5.991
1%	3.5956	31	0.1828	14.5765†	14.7593†	5.991
0.5%	4.4601	16	0.0312	25.9165†	25.9478†	5.991
0.1%	5.8811	5	0.7125	8.2298†	8.9424†	5.991
0.01%	-	-	-	-	-	-

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence are distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance of 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 18: HS - N=4037 - VaR estimate back-test

With a rolling-windows of 4037 observations (our maximum), we can see that the HS does not reject any of the left tail probabilities for the Kupiec test. However, concerning the Christoffersen test, the results remain unchanged. Nonetheless, it shows that with much more observations, the Historical Simulation yields to better results.

In our multiple Historical Simulation models, we observe that, even if the Christoffersen test are not conclusive, we are closer to non-rejection when p is equal to 0.1% compared with the other probabilities.

This lack of performance can be explained by 2 different aspects. First, as mentioned before, there is a problem of calculation for confidence intervals when the VaR is smaller than 1 in the windows size. Second, the empirical distribution for the HS Model is a step function, so it can bias our results.

4.2 Age-weighted

In this section we try to avoid the main issue of Historical Simulation (HS), namely equally weighted returns in the rolling window. In the standard HS, each return in the rolling window has the same weight. Whereas here, we attribute more weight for recent returns. We proceed as follows:

the weight w_i is :

$$w_i = \frac{\lambda^{i-1}(1-\lambda)}{(1-\lambda^N)} \tag{29}$$

λ is the the memory parameter. We choose $\lambda = 0.98$ as suggested in the FRM (Financial Risk Manager). We sort the subsamples and attribute each weight according to corresponding date for each return. Then, we compute the cumulative sum of the weights of the sorted subsamples. The VaR is the return when the cumulative sum is above θ, θ being the VaR level like 5%. The VaR is not directly computed at the VaR confidence level. Looking at the cumulative sum of the probability, we choose the VaR when the cumulative sum exceeds the confidence level of the VaR we consider. Hence, we have small underestimations of the risk measure. One might use a mean between the VaR estimate when it exceeds the confidence level and the previous estimation or only consider the last value before it exceeds the confidence level. The latter tends to overestimate the VaR. We apply for different rolling window sizes.

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.6925	186	2.1622	24.2016†	26.3638†	5.991
1%	2.8527	53†	9.8248	3.4979	13.3227†	5.991
0.5%	3.1822	40†	23.3992	6.2993†	29.6985†	5.991
0.1%	-	-	-	-	-	-
0.01%	-	-	-	-	-	-

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence are distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance of 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 19: HS Age-weighted - N=250 - VaR estimate back-test

In Table 19, we observe that the Age-weighted approach for the HS improves the results that we have with a standard Historical Simulation with 250 observations slightly. Indeed, we don't reject the Kupiec test for p=5%, however we do for p=1% and p=0.5% as with the standard Historical Simulation and N=250. No improvement for the Christoffersen test, we still reject the null for every p.

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.6934	185	1.9423	24.6513†	26.5936†	5.991
1%	2.8568	52†	8.9106	3.6726	12.5832†	5.991
0.5%	3.1889	40†	23.3992	6.2993†	29.6985†	5.991
0.1%	-	-	-	-	-	-
0.01%	-	-	-	-	-	-

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence are distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance of 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 20: HS Age-weighted - N=500 - VaR estimate back-test

Same as Table 19. We do not reject the Kupiec test for p=5%, however we do for p=1% and p=0.5%.

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.6934	185	1.9422	24.6513†	26.5936†	5.991
1%	2.8568	52†	8.9106	3.6726	12.5832†	5.991
0.5%	3.1889	40†	23.3992	6.2993†	29.6985†	5.991
0.1%	3.7746	24†	53.4309	1.9195	55.3504†	5.991
0.01%	-	-	-	-	-	-

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence are distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance of 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 21: HS Age-weighted - N=1000 - VaR estimate back-test

In table 21, the results are the same than for the Age-weighted with 250 and 500 observations : a non-rejection of the Kupiec test with a probability of 5% and all the other tests rejected.

As regards the VaR for N=1000, graphically, see Figure 6 in the Appendix, we can see that the smaller is the significance level, the longer the changes/shocks last. That is, the changes in VaR value are more persistent over time and take longer to go back to its pre-shock value.

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.6934	185	1.9422	24.6513†	26.5936†	5.991
1%	2.8568	52†	8.9106	3.6726	12.5832†	5.991
0.5%	3.1889	40†	23.3992	6.2993†	29.6985†	5.991
0.1%	3.7746	24†	53.4309	1.9195	55.3504†	5.991
0.01%	-	-	-	-	-	-

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence are distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance of 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 22: HS Age-weighted - N=4037 - VaR estimate back-test

Same as table 19, 20 and 21. Nonetheless, in the case with 4037 observations, the standard Historical Simulation has better results because all the Kupiec tests are not rejected compared to the Age-weighted simulation which is not rejected only for the left tail probability of $p = 5\%$.

As regards the VaR, we can say that the Age-weighted Historical Simulation with a subsample of 250 observations is better than the standard one because for any given p, it has a lower VaR than the standard HS. For an N=500, the VaR is the same for both methods for p=5% but it is smaller in the case of the Age-weighted method for p=1% and p=0.5%. For N=1000, we can clearly see a significant improvement in the VaR value between the standard and the Age-weighted HS, the VaR is lower for every p in the case of the Age-weighted method. To sum up, the Historical Simulation method doesn't give any interesting results except for the standard Historical Simulation with 4037 simulation which is the best since it doesn't enable to reject the Kupiec test for any confidence level.

To conclude the part of Age-weighted models, we observe that, since we have 500 observations or more, the computations give exactly the same results. It might come from the robustness of the methodology. By over weighting the recent returns and under weighting the old returns, increasing the rolling window has no effect when the weight is decreasing as the return is older.

5 Extreme Value Theory

In the two following sections, we use the Extreme Value Theory (EVT) in order to obtain an appropriate quantile to compute the VaR. We use the two approaches, namely the extreme one and the tail one. In the risk management point of view, we are essentially concern about the tails event and, in particular, the negative tails'. The EVT focuses on the tails of the distribution giving a smooth function for the tail. Please notice that we do not apply directly the EVT to the returns. First, we filter the returns by withdrawing the unconditional mean and the volatility with a GARCH in order to retrieve the the innovation process. It is useful to go in such a way because the EVT requires iid process which is not the case for returns but it should be so for z_t. The extreme approach need iidness across the subsample while the tail approach need it for the whole sample. The only discretion we have in the model is the ability to choose the subsample window for the Extreme approach and choosing the number of excess return for the tail approach.

5.1 Extreme approach

In this section, we use the Extreme approach to describe the tail of the distribution. As model, we consider the one describes in equation 30. We use the unconditional mean and a GARCH(1,1) process for the conditional variance. This allows us to filter the return process to get the innovation process which is iid as required by the EVT. To estimate GARCH parameters, we use the QMLE approach with the normal distribution. This gives us consistent parameters but not efficient. Nevertheless, using the normal avoid having inconsistency due to misspecification of the distribution.

$$
\begin{aligned}
&r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad z_t \sim iid(0,1) \\
&\sigma_t^2 = \omega + \alpha\varepsilon_{t-1} + \beta\sigma_{t-1}^2 \\
&z_t = \frac{\varepsilon_t}{\sigma_t}
\end{aligned} \tag{30}
$$

According to the results below, we always reject the independence test which suggests that the z_t process might not be iid. We perform a Ljung box test for the innovation to verify if, at least, they are not correlated. We have the same innovation process for the tail approach, hence this result also holds for the next section. We report the result in Figure 23.

Lags	Ljung-Box	pValue
1	3.4426	0.0635
2	3.4474	0.1784
3	4.5915	0.2043
4	8.1403	0.0866
5	10.3935	0.0648
6	10.4762	0.1060
7	10.6171	0.1562
8	10.6401	0.2229
9	11.3764	0.2507
10	11.5004	0.3199

Table 23: Autocorrelation in z_t

In the absence of the other independency test at hand, we performed a Ljung-Box test, and we can see that there is no evidence of auto-correlation in the innovation process as we never reject the null hypothesis at 5% and lower. We do not interpret that as independence as it is not a sufficient condition under non-normality. The same result applies also for the Tail approach, hence we do not report the Table 23 in the next subsection.

Once we get the vector of z_t for the estimation window, we split the estimation sample into subsamples. The size of each subsample is 60 days, corresponding to a quarter. On each subsample, we compute the maximum value and end up with a new time-series of maximum for the estimation window M_t. The length of this vector of maximum is 67. We fit the Generalized Extreme Value Distribution and get the tail index ξ, the scale parameter ψ and the location parameter μ. Then, we are able to compute the quantile q_θ^z for different levels of θ using the following equation:

$$q_\theta^z = \hat{\mu} + \frac{\hat{\psi}}{\hat{\xi}} \left[(-N * log\,(\theta))^{-\xi} - 1 \right] \tag{31}$$

Finally, we compute the the VaR_{t+1} with the prediction for the σ_{t+1} and μ using the same formula as before.

To estimate the GARCH model, we basically use the same as in Dynamic models (GARCH)- Normality. The parameters are significant but with lower t-stats than if we had used MLE. To see the parameters, please refer to Table 9.

Parameter	Estimate	t-stat
ξ	0.3406	2.9584
Ψ	0.5096	8.2226
μ	2.1824	30.3928

Table 24: GEV - Distribution Parameters estimated

In Table 24, the most important parameter is the ξ, the tail index. It indicates the shape of the GEV distribution. As we have financial returns, we expect to find a value for ξ at least larger than or equal to 0 as it corresponds to possible distribution for the innovation process, namely the normal or log normal distribution in the case of $\xi = 0$ or a student-t or a skewed-t if $\xi > 0$. The test statistic shows that the ξ is significant. The value of ξ is higher than 0 indicating that the extreme values are in the domain of attraction of a Fréchet, which is typical for financial returns. Hence, the tail is non-normal. We can see this graphically in Figure 8.

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.8207	155	0.9518	15.2961†	16.2479†	5.991
1%	2.6207	49†	6.4052	7.6325†	14.0378†	5.991
0.5%	3.1266	24	0.8098	6.4390†	9.2488†	5.991
0.1%	4.9045	3	0.0365	10.5926†	10.6291†	5.991
0.01%	10.1266	0	0.6686	NaN	NaN	5.991

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence are distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance of 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 25: Extreme approach - VaR estimate back-test

Unlike the HS, we are able to have good estimates of each confidence level of the VaR,in particular for very high level such as 0.01%.

Regarding the back test of the Extreme approach with a GARCH, the unconditional test gives good results except for the VaR at 1%. The failure rate is too high compared to the theoretical predictions. This is the only level where we reject the null hypothesis. In our sense, several issues can generate this result. It could be due to the fact that we have too short samples of extreme or it might come from the special shape of the Fréchet distribution. We expect good results from this model as it is one of the most advanced methodology. We can see graphically the evolution of the VaR at Figure 9.

Regarding the conditional coverage test, we always reject the null hypothesis, which is disappointing. Nevertheless, it is one of those that has the lowest value for the test statistic, we are not very far from the critical value. We could be in the case where we reject the null while the null is true (type 1 error). The rejection means that the violation of the VaR are too close to each other. We think that it might come from the period of the GFC where the disturbance last much longer than event such as COVID-19 crisis. Even if the GARCH adjusts relatively well the level of the VaR, we think that another problem arises from the fact that our methodology is not completely dynamic because we use parameters estimated in a static way. We can see this graphically in Figure 10. We observe that the occurrences of the VaR are particularly close to each other in certain periods notably in 2008-2009, 2011-2012 and 2020. The small spikes are very close during these periods. This result is also applicable to the Tail approach.

5.2 Tail approach

In this section we use the other approach of EVT. We expect to have similar results as in the previous one because the methodology is quite similar. Basically, we use the same assumption as above and same process as in equation 26. For GARCH parameter, please refer to Table 9. Results about auto-correlation are in Table 23.

Once we have the z_t innovations process, we sort the series. We define an upper threshold q corresponding approximately to 10% of the data in the estimation window. We have $N_u = 400$ and $T = 4037$. We define u, the threshold corresponding to the qth value. We collect the subsample corresponding to the value above q and subtract u to get the exceedances. Finally, we fit the Generalized Pareto Distribution using the exceedances. We compute the corresponding quantile q_θ^z as follows :

$$q_\theta^z = u + \frac{\hat{\psi}}{\hat{\xi}} \left[\left(\frac{T}{N_u} * (1-\theta) \right)^{-\xi} - 1 \right] \tag{32}$$

We can, then compute the VaR using the new quantile.

Parameter	Estimate	t-stat
ξ	0.0830	2.0336
Ψ	0.5782	15.5229
u	1.2430	-
q	400	-

Table 26: GPD - Distribution Parameters estimated

First, we report a positive u as we use the loss function. The EVT tells us that we should find the same parameter ξ as with the GEV. Using the GPD, we find a ξ that is close to zero suggesting that the the domain of attraction of the tail is such a normal or a log normal. The result is a bit disappointing. We think that is comes from the fact that due to the length of our sample, we do not have sufficiently a long subsamples to estimate correctly the parameters. Hence, it might be the case that we have central returns contaminating the process. Obviously, it would impact the estimation of the VaR.

p	Average VaR estimate	Failure rate	LR_{unc}	LR_{ind}	LR_{cc}	Critical value (5%)
5%	1.6707	194†	4.2405	20.7840†	25.0245†	5.991
1%	2.7858	39	0.8898	10.9750†	11.8648†	5.991
0.5%	3.3191	18	0.0968	2.9540	3.0508	5.991
0.1%	4.6980	3	0.0365	10.5926†	10.6291†	5.991
0.01%	7.0721	1	0.8602	0.0006	0.8608	5.991

The "†" sign indicates that the test is rejected. Please note that the test of Kupiec and the test of independence are distributed as a $\chi^2(1)$ and the test of conditional coverage is distributed as a $\chi^2(2)$. We perform a test with a level of significance of 5%. It means that the critical value of the $\chi^2(1)$ is equal to 3.84 and the critical value of the $\chi^2(2)$ is equal to 5.991. We report the number of failures in the table. However, in order to compute the failure rate we have to divide the number of failures by 3343 which is the evaluation window's size.

Table 27: Tail approach - VaR estimate back-test

Firstly, practitioners consider this approach preferable over the extreme one in terms of data as it considers the whole sample above the threshold rather than the maximum in each subsample. We can observe the evolution of the VaR at Figure 9.

Regarding the unconditional test, we reject the null hypothesis for the VaR at 5% confidence level but not at 1%. We have a relatively high number of exceedances at 5% compare to the GEV. However, it is not very far from the non-rejection zone. This might be due to the shape of the distribution that we can see in Figure 12.

The conditional coverage gives better result than the GEV. Indeed, the model does not reject the test at 0.5% and 0.01%. However, for the ones that we reject, the test statistics are lower for the GEV than for the GPD. We think that the rejection are due to the same reason as for the GPD. There is too many exceedances during the turbulent crisis.

It is hard to see a model that stands out from the other by simply comparing the results. Both lead to similar results in term of rejection. We are not able to identify a model that is always closer to the non-rejection zone.

Conf.level	GEV	GPD	$\mathcal{N}$
q_{θ}^{z}	2.4637	2.7036	2.3263

Table 28: Quantile for VaR 1%

Finally, the Table 28 shows the impact of quantile on the estimation of the VaR using the EVT instead of the Normal distribution. At 1%, the impact is relatively important and it increases as we increase the confidence level of the VaR. It shows how we can undervalue the VaR using too simple model. We can also see this difference graphically in Figure 12. We also observe that the distribution of the Generalized Pareto is very different from the Fréchet. Indeed, the GPD decreases much faster and explains the different results for the quantile.

6 Discussion

6.1 Models analysis

In this section we compare our model between each other. We try to select the best one among each methodology. Namely, the variance method, the historical simulation method and the extreme approach.

First of all, using the parametric method (variance method), the best model is the dynamic Student-t distribution. Indeed, the unconditional test is never rejected for every left tail probabilities p. It only rejects the independence test (hence the conditional coverage test) for $p = 5\%, 1\%, 0.1\%$. Therefore, it is fundamental to take into account the fat-tails of the returns distribution to have a good risk management model. Initially, we expected the skewed-t to give better results but it gives too low VaR estimates because of the Skewness of the distribution. This is graphically showed in Figure 13. We can therefore conclude that there is a trade-off between capturing Skewness and capturing excess Kurtosis. In fact, by taking care of the asymmetry of the returns, we capture less excess positive Kurtosis (we cannot attain every values with a skewed-t distribution). We know that in a risk management point of view, it is more important to take into account the fat-tails of the distribution than the asymmetry of data.

Concerning the historical simulation method (non-parametric method), just the age-weighted method stands out from the standard Historical Simulation approach. Indeed, we reject the conditional test for all left tail probabilities p whatever the number of observations we have, and also the unconditional test for every p except for p=5%, except for the Historical Simulation with N=4037 observations where we do not reject any of the unconditional tests. It is our best model for the non-parametric method. In principle, Historical Simulation should yield to interesting results because it takes the non-normality into account from a sample of the true distribution. Moreover, the age-weighted methodology should give particular good estimations of the VaR in turbulent periods. It also resolves the problem of ghost effect regarding the standard historical simulation. Nevertheless, for most of these non-parametric techniques, they give unsatisfying results compared to the more advanced one.

Finally, regarding the extreme value theory techniques, both gives similar results. The tail approach seems to be preferable over the extreme approach as it consider the whole sample when the extreme use only one value over sub sample. However, the extreme approach is more likely to give a iid process. In our case this is not really a problem as we use the innovation process which we assume to be iid. Hence, we think that the tail approach with a GARCH is preferable over the extreme approach. The tail approach is a bit less conservative than the others but gives good result.

To sum up, we can stipulate that in our study, the dynamic Student-t distribution from the parametric approach performs the best among every other techniques.

6.2 Limitation & improvements

In terms of improvement of the model, we can argue that we might consider to use the tail approach with a GARCH(1,1) with the Skewed t distribution or at least Student t distribution in order to capture all the properties of the innovation process. By mixing both approach, we expected to have even better results. We still have to keep in mind that if we make an assumption on the distribution of the innovation process and that we are wrong, we would have inconsistent estimates of the parameters. Which is not the case using the normal distribution and the QMLE approach.

An other improvement could be to consider instead of a GARCH process for the volatility a stochastic model. Indeed, they are much more flexible than the former. We could also consider extension of the GARCH such as GJR model in order to capture the asymmetry in volatility process. An others possibility is to use a stochastic volatility model which are more flexible but more intense in computation.

Finally, regarding the limitation, we use a particular period for estimate parameter involving the global financial crisis, while back testing for a bullish and relatively calm period. It might arise some problems in term of conservatism, while our parameter can make use overestimating the VaR for the period we back test impacting directly the statistical test such as Kupiec. Moreover, in a practical point of view, splitting the sample over two periods, estimate parameters with the first sample and use them in the other sample might be not the best approach. Use a rolling window for each new VaR estimate allow to adjust the parameter estimate and consider actual data. The use of a rolling window imply to choose a size for this one. It could have some consequence on the estimate. The size has to be choose in a way that take extreme event in the sample while not take too old data that might be completely disconnect with actual time.

We see also a limitation with the extreme value theory. Indeed, it requires a huge data set for the quantile to be consistent. We think that our data set is not large enough for this methodology.

The last limitation is the risk measure itself. Indeed, the VaR is not a coherent risk measure in all the case. Consequently, it is more appropriate to use the expected shortfall which is a more coherent risk measure.

7 Conclusion

Throughout this study, we realised the importance of implementing an appropriate risk management model in order to achieve a balance between compliance with banking regulations and the opportunity to generate profits. Indeed, we have found that if we use a model that is too aggressive, the regulator may introduce punitive taxes and fines and therefore incur a cost to the financial institution. Conversely, an overly conservative model would imply a misallocation of financial resources. Indeed, if the financial institution keeps a much higher capital requirement than necessary, this will have an impact on the cost of capital. In other words, the institution will not be able to take advantage of growth opportunities. It is therefore fundamental to have a model that neither underestimates nor overestimates market risk.

We saw that it is very important to capture all the characteristics of the financial data in order to build the most adequate model possible. Firstly, it is important to make an adequate assumption about the distribution of returns that takes into account the excess positive Kurtosis implied by the data. Secondly, it is also fundamental to consider the change in volatility over time. This also implies the notion of fat-tails in the distribution of our data. In addition, with a GARCH model, we can build a model that can adapt to the change in volatility in the market in order to have the capital required to deal with it. Lately, we know that in a risk management model, we focus principally on extreme events that may occur. To this end, we have presented the Extreme approach and tail approach from the EVT in order to characterise the tail of the distribution.

To conclude, we can elaborate further on the impact of the recent COVID-19 crisis on our risk management models. Indeed, we have seen the importance of taking these events into account in the data in order to have a correct and adequate model. If we were to take no extreme events in our data, it would therefore be very difficult to have a good risk management model if one of these adverse events occurs in the future. It is therefore fundamental to implement a lot of data in our model in order to take into account as many adverse events as possible that may occur in the future (like the 2008 or COVID-19 crisis). We have seen in this crisis that financial institutions are strong and have set aside enough liquidity and capital to continue their operations adequately. The reason for this is due to the very strict regulations issued by authorities after the financial crisis in 2007, when banks were in very great difficulty. The soundness of banks is fundamental to ensure the sustainability, to recovery of the global economy and to avoid systemic risk.

8 Appendix

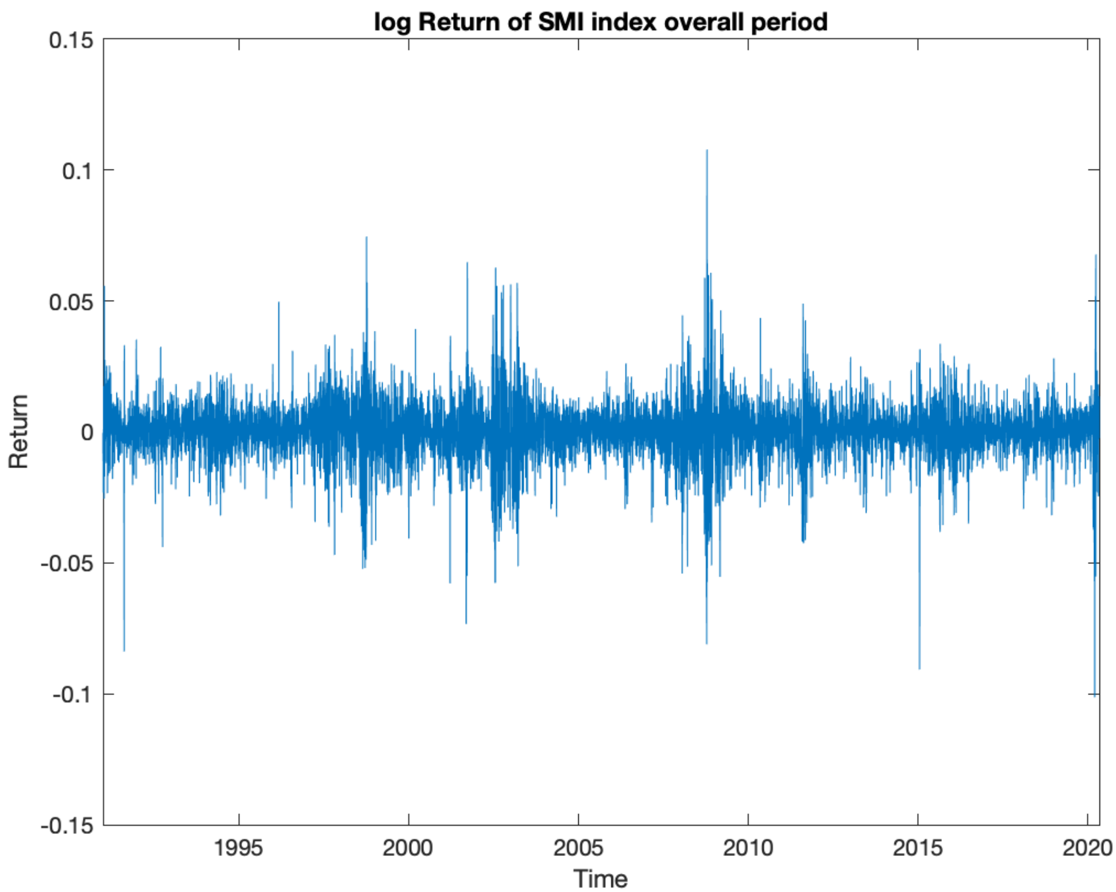

Figure 1: Log return SMI (overall period)

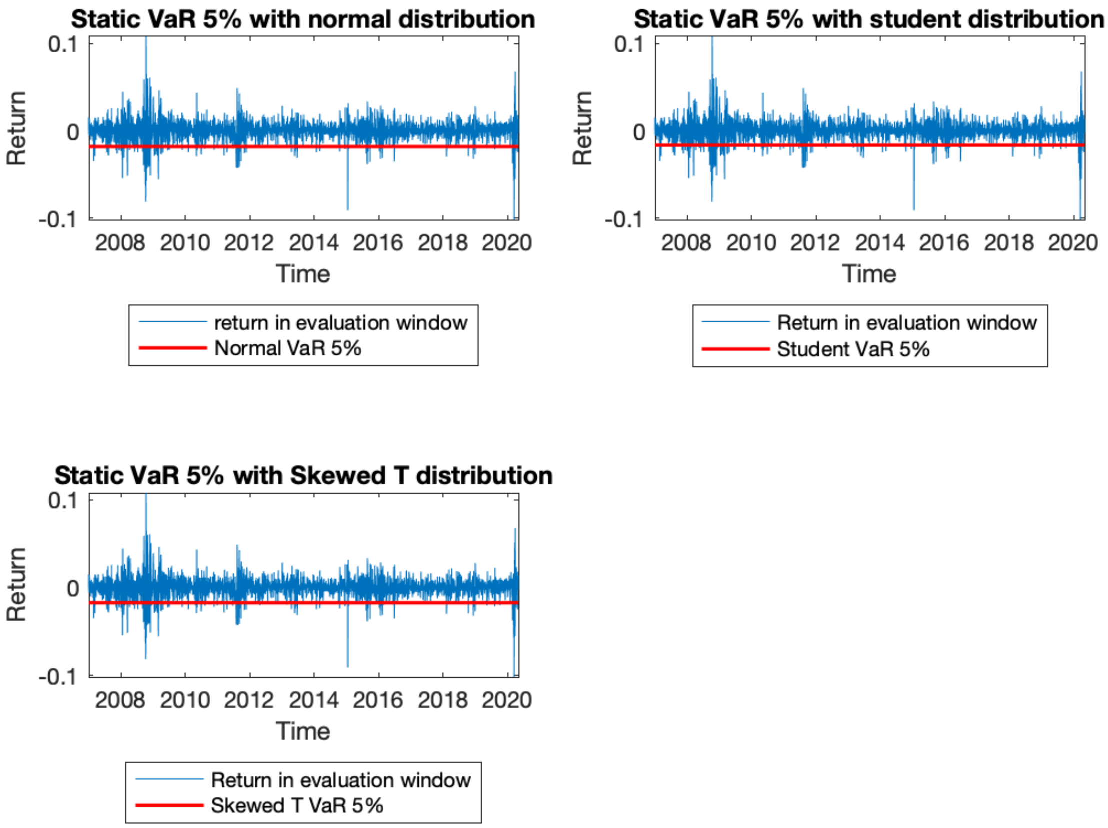

Figure 2: Static VaR (5%)

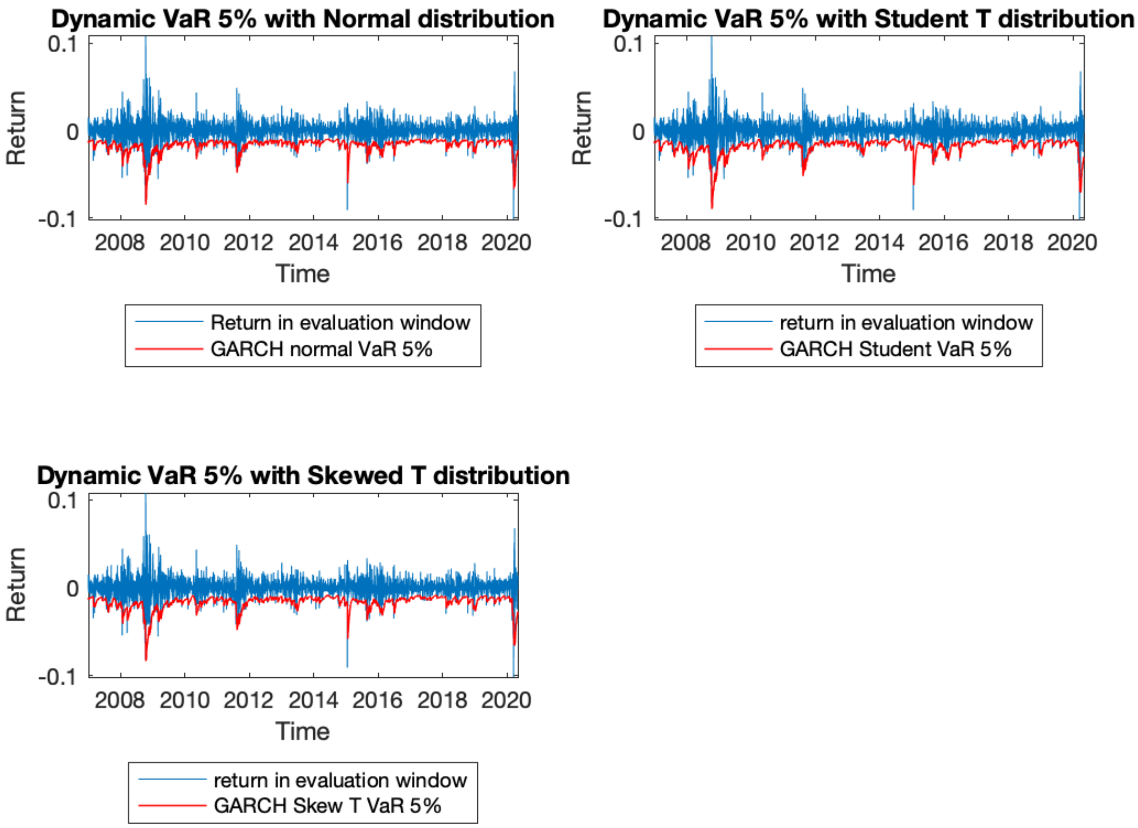

Figure 3: Dynamic VaR (5%)

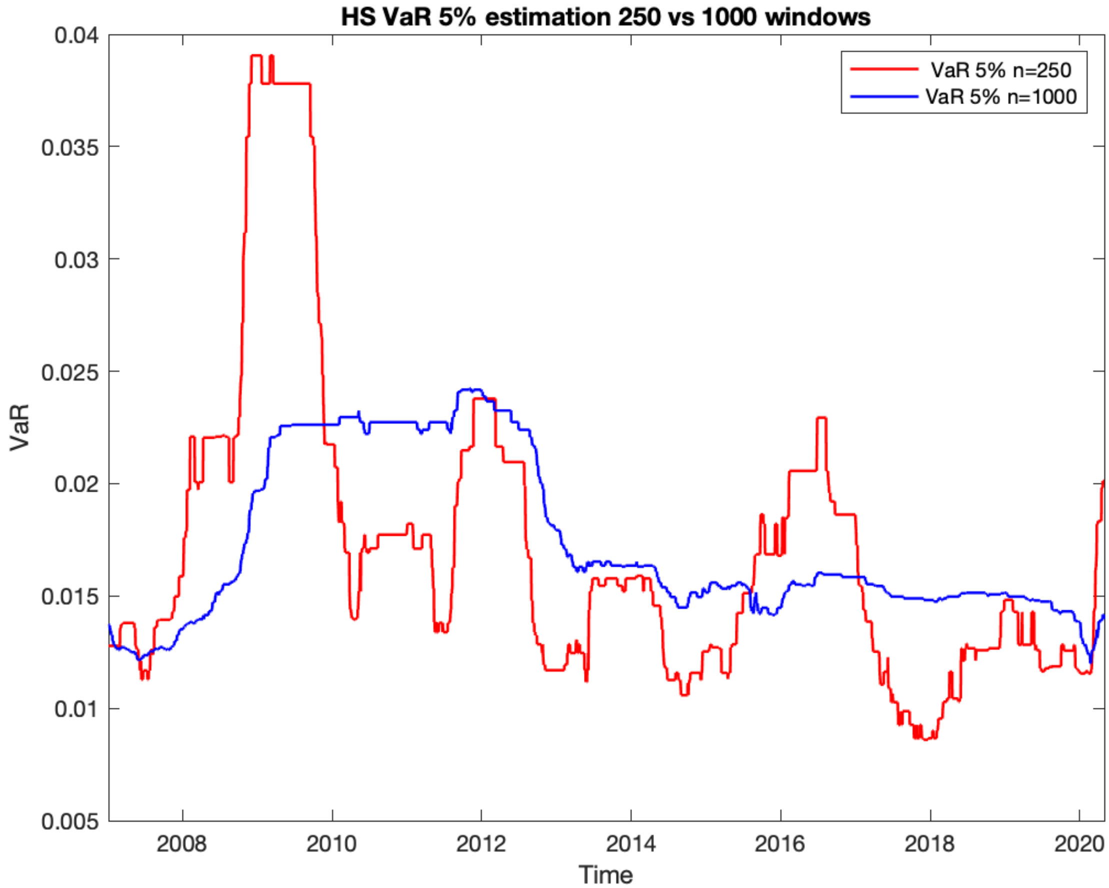

Figure 4: HS - VaR prediction using HS

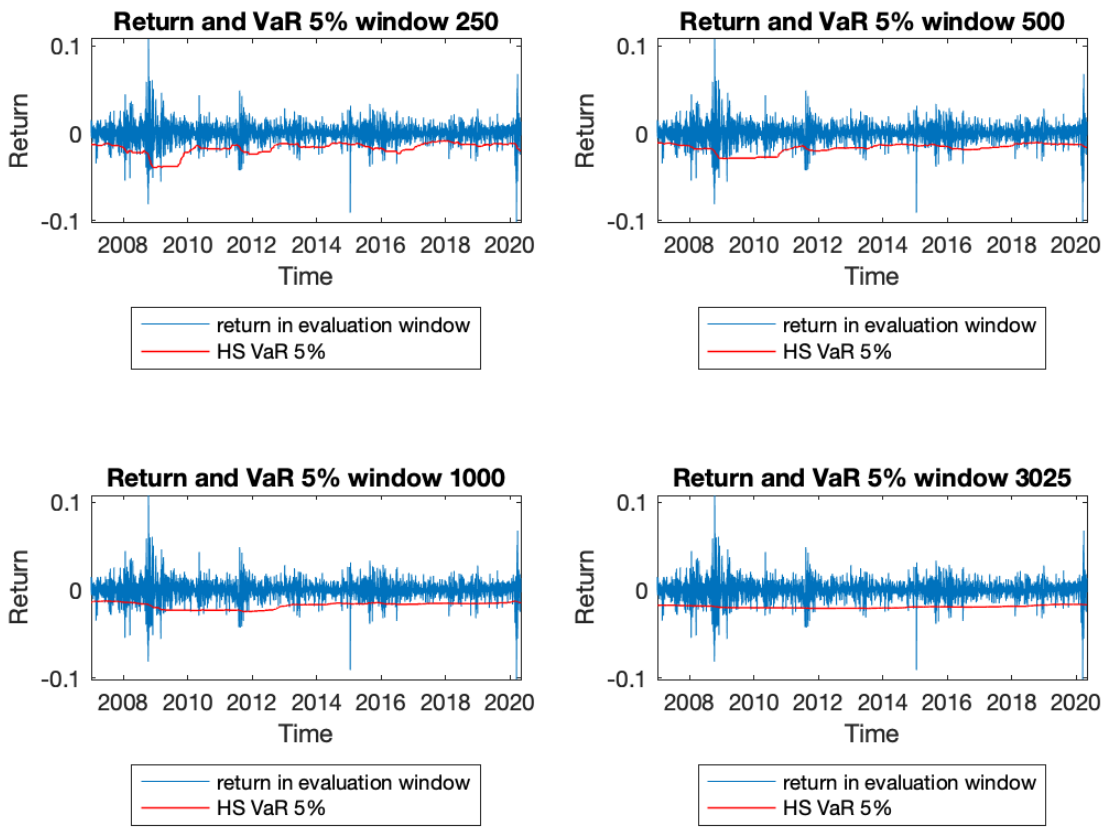

Figure 5: HS VaR 5%

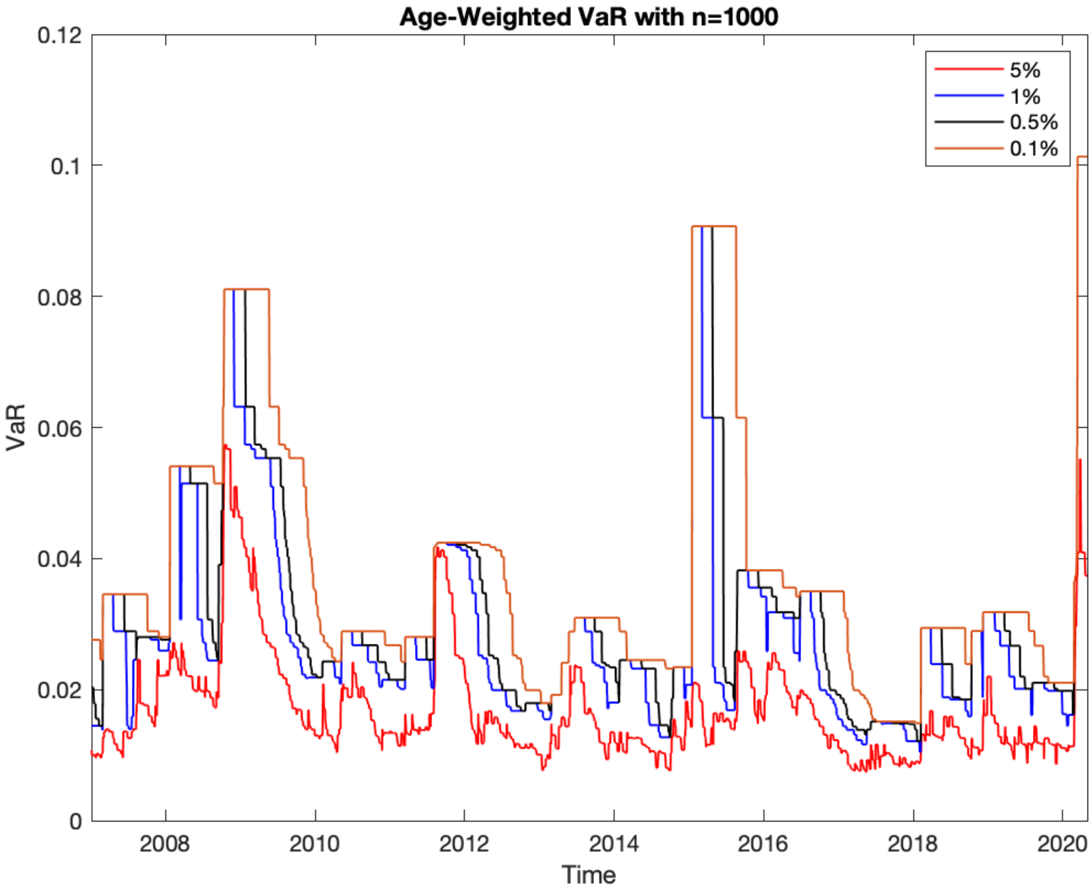

Figure 6: Age-Weighted - VaR prediction using HS age-weighted

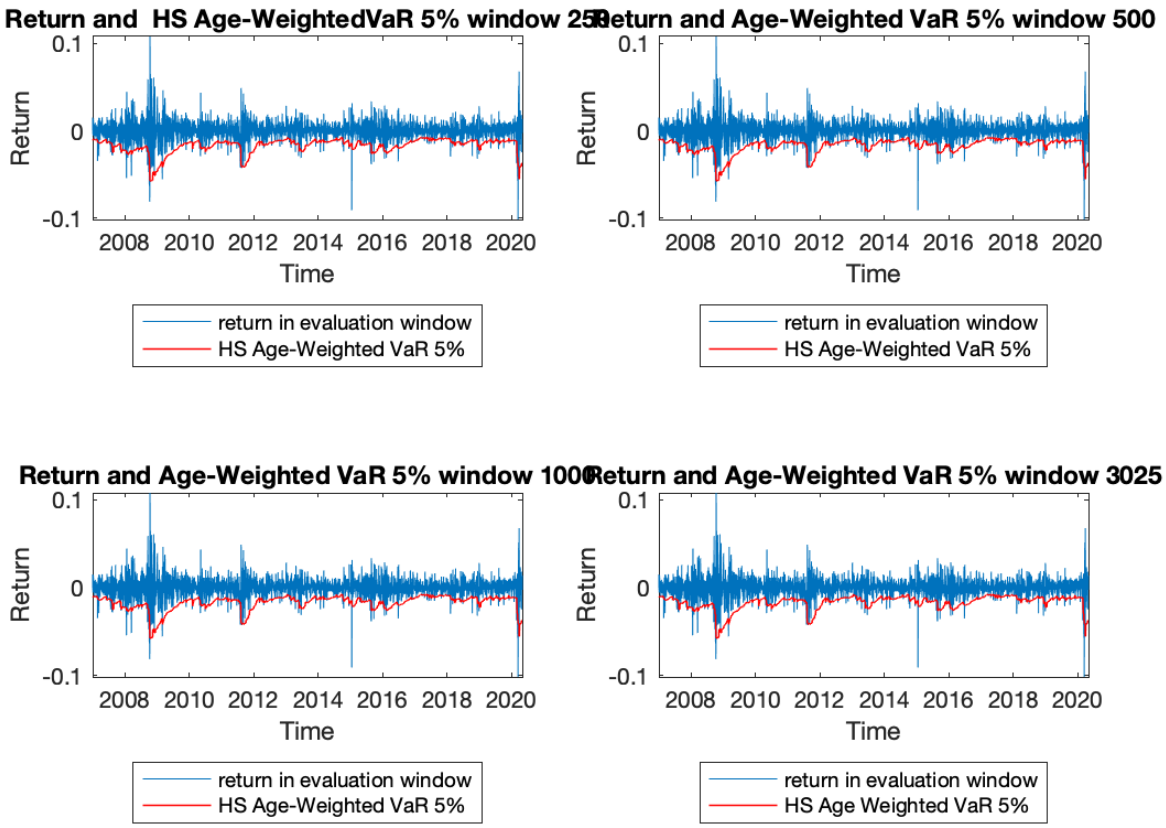

Figure 7: Age-Weighted VaR 5%

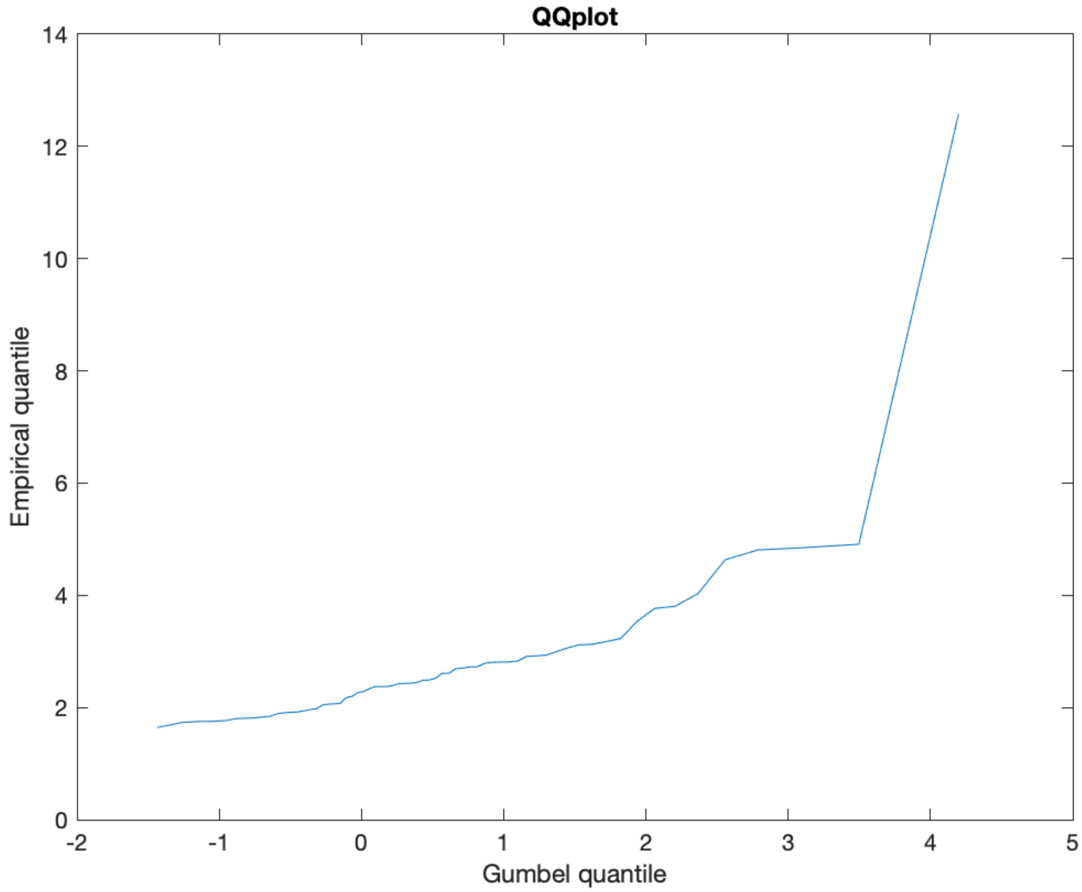

Figure 8: QQPlot - Extreme approach

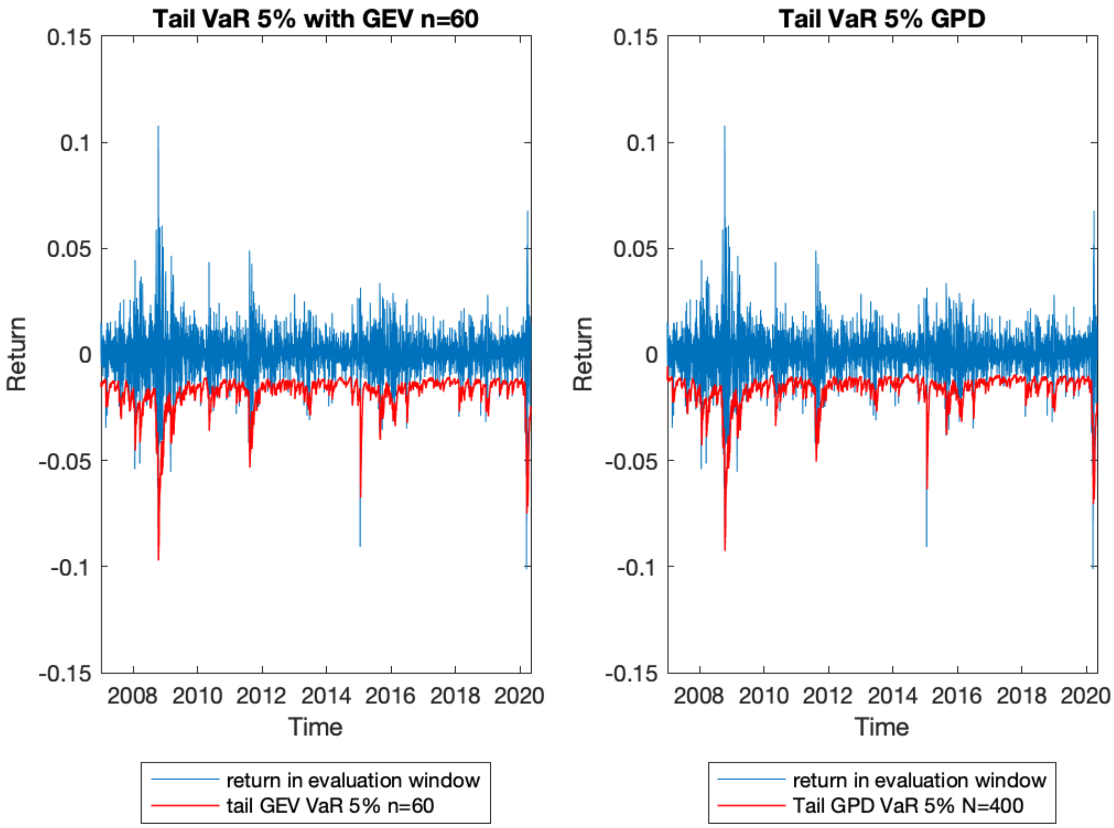

Figure 9: EVT VaR(5%)

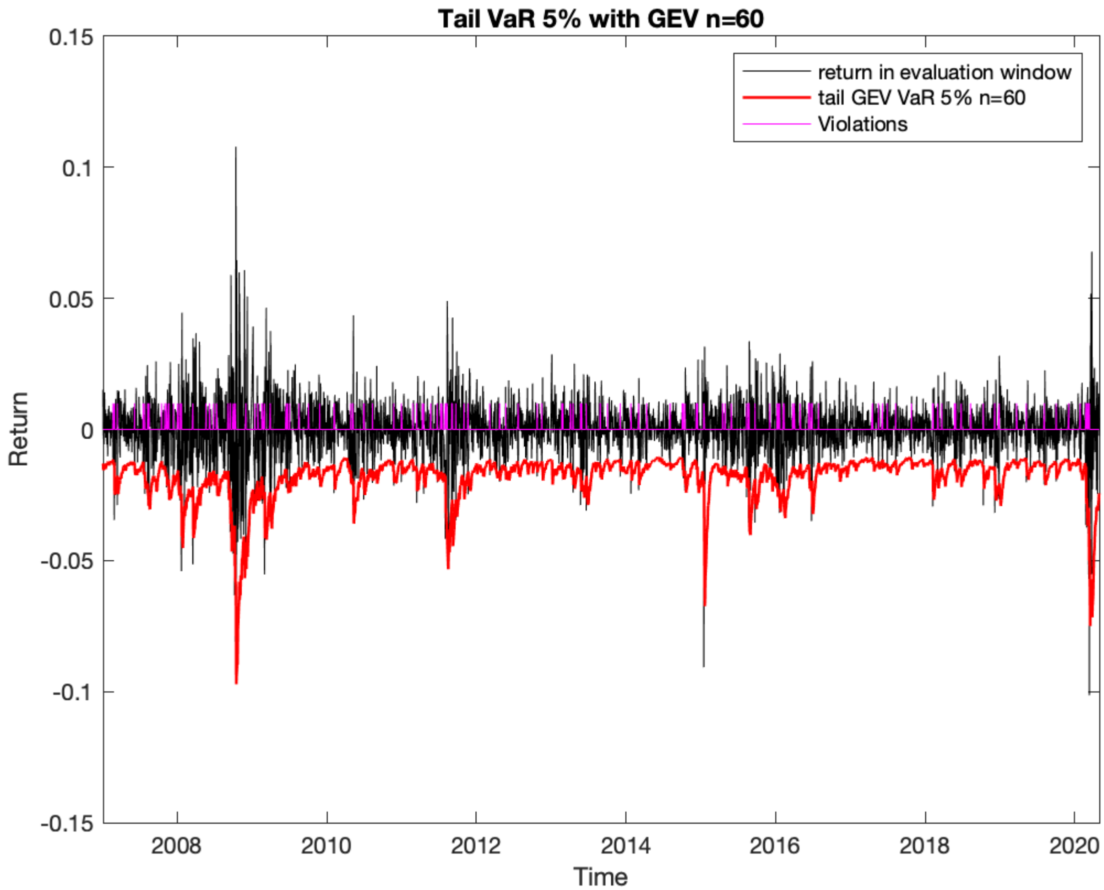

Figure 10: EVT GEV VaR(5%) with violations

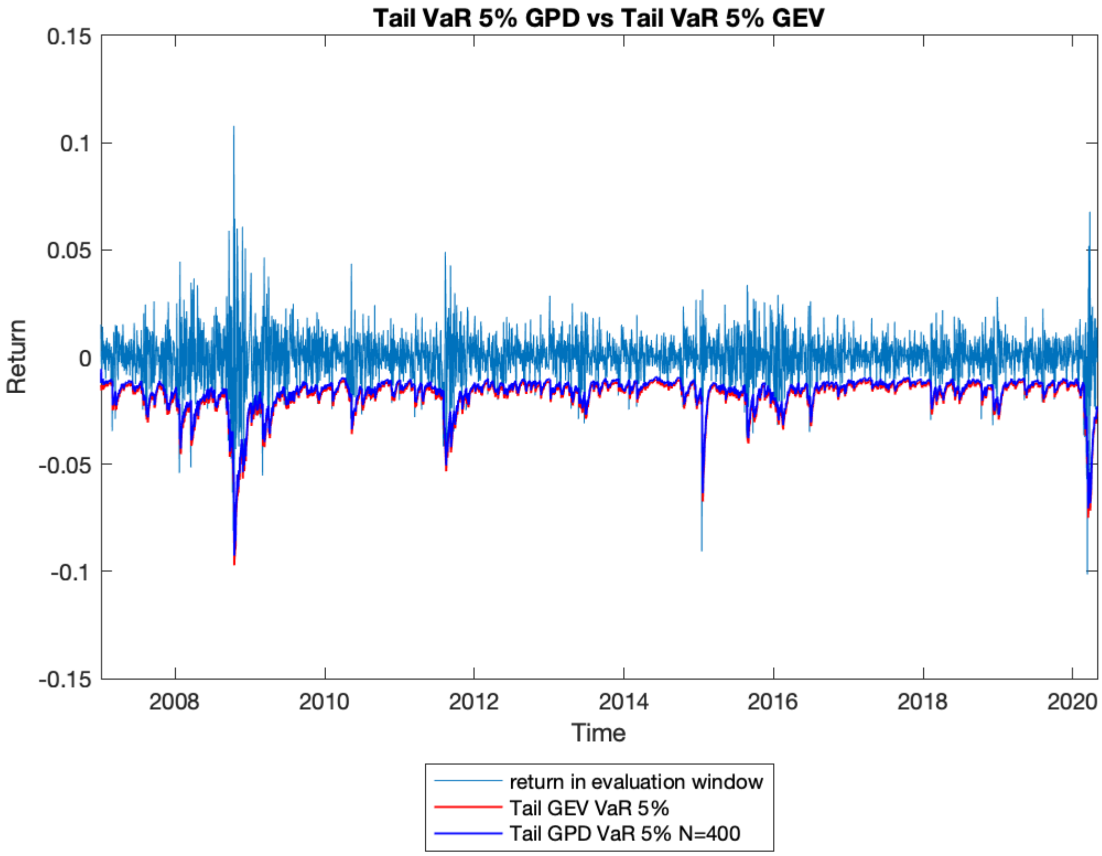

Figure 11: Comparaison GEV and GPD VaR(5%)

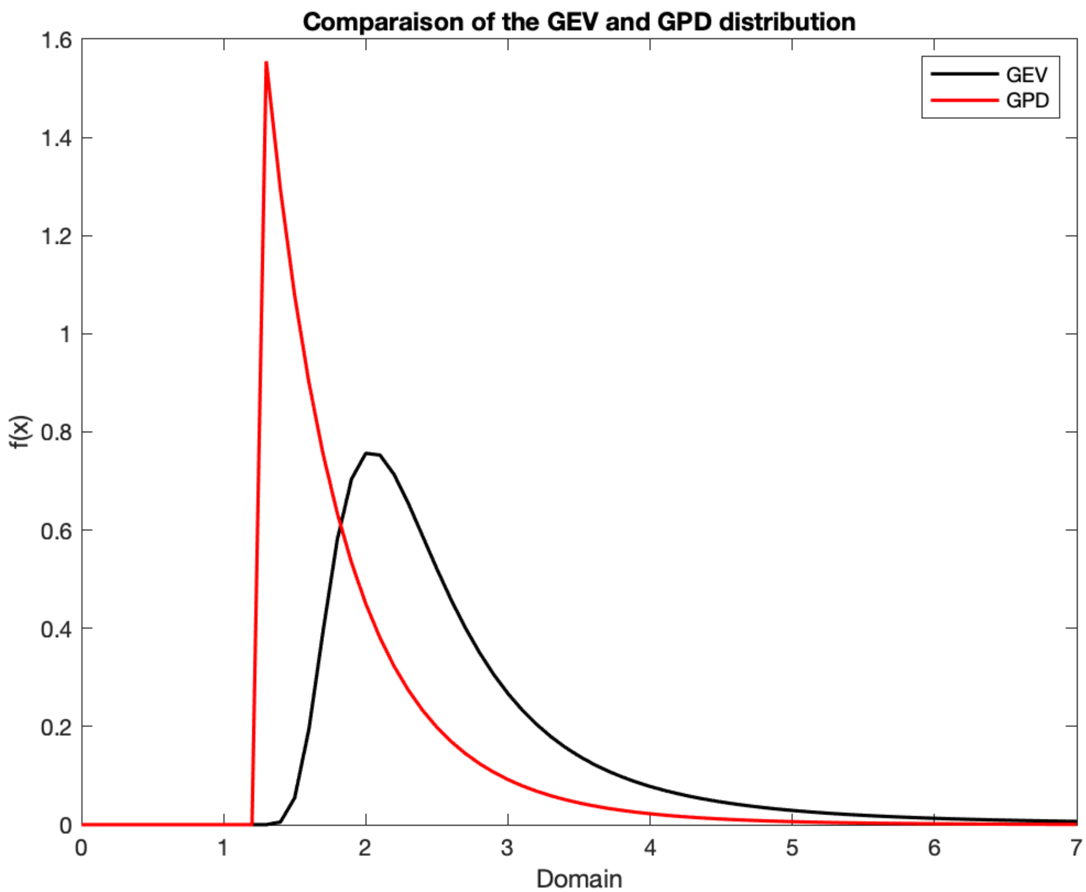

Figure 12: Distribution of the GEV and the GPD)

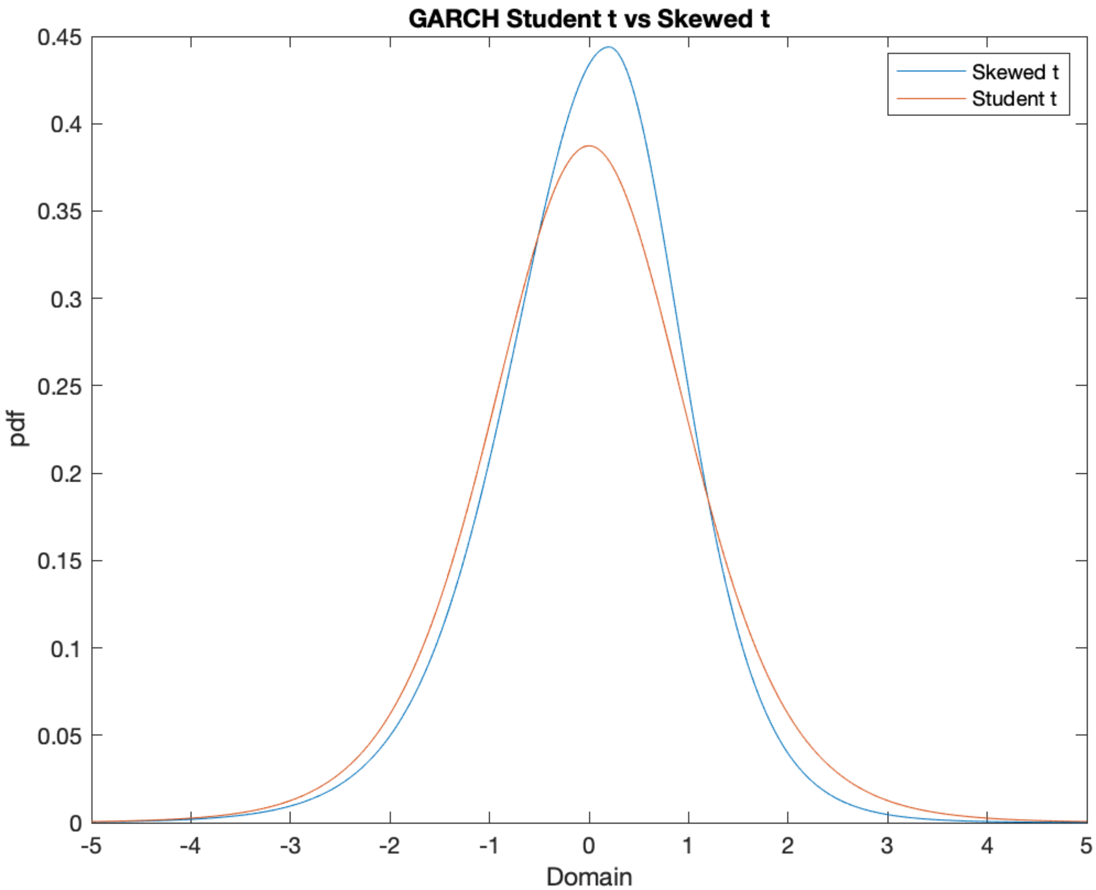

Figure 13: Distribution z_t in GARCH Skewed t vs Student t

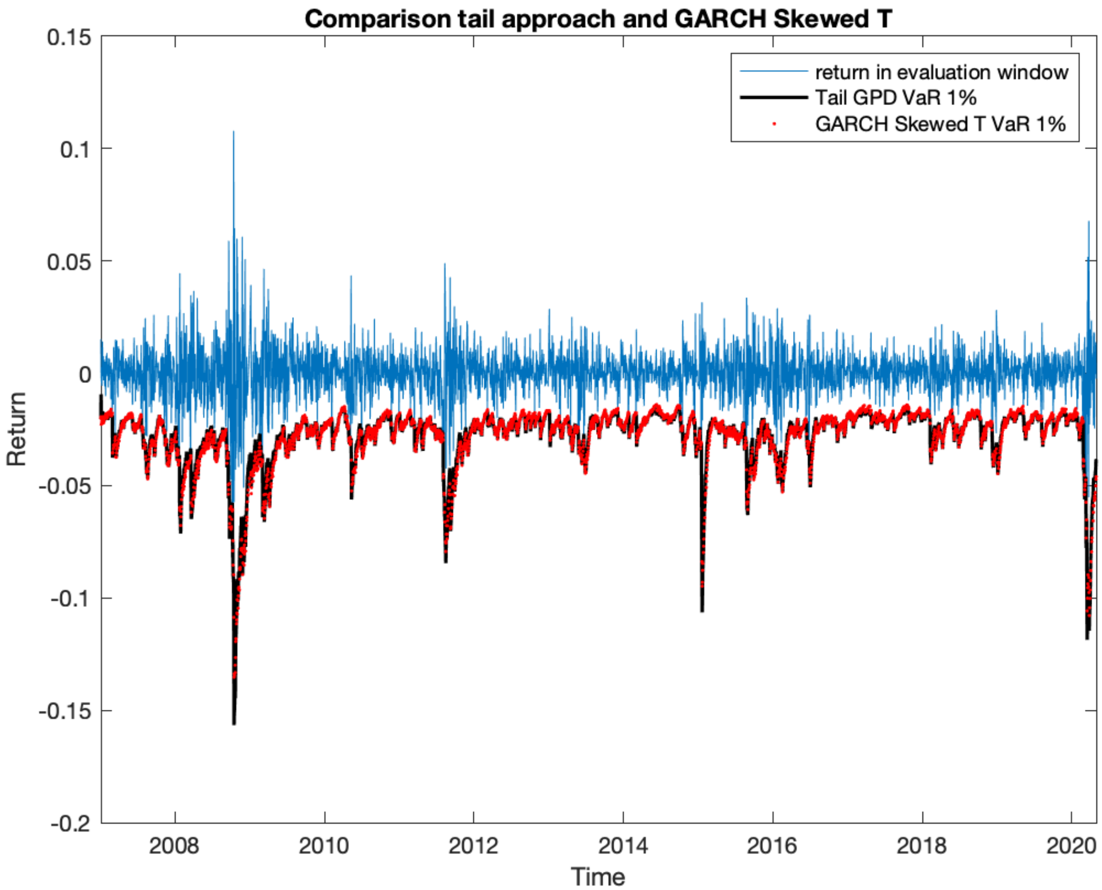

Figure 14: Comparison Tail approach and GARCH Skewed T VaR(1%)

References

[1] Rob van den Goorbergh, Peter Vlaar. *Value-at-Risk Analysis of Stock Returns Historical Simulation, Variance Techniques or Tail Index Estimation?*. 1999.

[2] Peter F. Christoffersen. *Evaluating Interval Forecasts*. 1998.

[3] Keith Kuester, Stefan Mittnik, Marc S. Paolella. *Value-at-Risk Prediction: A Comparison of Alternative Strategies*. 2006.

[4] Alexander J. McNeil, Rudiger Frey. *Estimation of tail-related risk measures for heteroskedastic financial time series: an extreme value approach*. 2000.

[5] Robert F. Engle. *Autoregressive Conditional Heteroskedasticity With Estimates Of The Variance Of United Kingdom Inflation*. 1982.

[6] Tim BOLLERSLEV. *Generalized Autoregressive Conditional Heteroskedasticity*. 1986.

[7] Yahoo Finance *Swiss Market Index*. Data extraction 2020
`https://finance.yahoo.com/quote/%5ESSMI/history/`