Financial Risk
Technical Project 1

Ramin Maghsood and Christos Saltapidas

May 6, 2018

1 Introduction

In this project we are going to analyze a financial time series of American multinational investment bank, the Goldman Sachs. We consider a time period of the last year (April 2017-April 2018) in our analysis and compare it with a second time period (August 2008-August 2009) where the financial crash of September 2008 is included. The comparison between these two periods would be important, since the difference and similarities of our results will direct us to be able to comment on the risk.

2 Aim

The goal of our project is to analyze the features of Value at Risk (VaR) in two different time periods with the use of three methods: Assuming Normality, the Block Maxima method (BM) and the Peaks over Thresholds method (PoT).

3 The Data

We downloaded a historical daily closing prices of Goldman's stock for 1 trading year for each period from Yahoo Finance. Then we saved them in two different vectors in MATLAB so we could start the analysis.

It is known that when using stock prices as data, we should work with the log-returns. We had some problems and some strange results so, we chose to work with just the differences between daily prices. Data are almost one year of trading days (251 days) of Goldman's daily closing market price per share, so the difference is a vector of 250 elements (it is always better to have a nice number, that's why we have chosen 251 days). Then we multiply that daily difference by -1 to get the daily loss. A negative loss of course, means return.

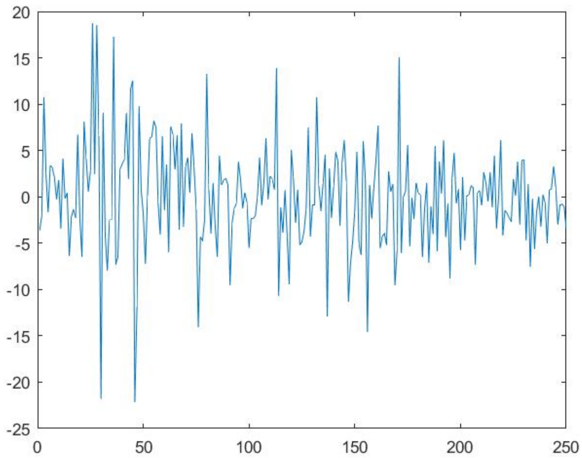

Figure 1: 2008 Goldman Sachs Daily Losses

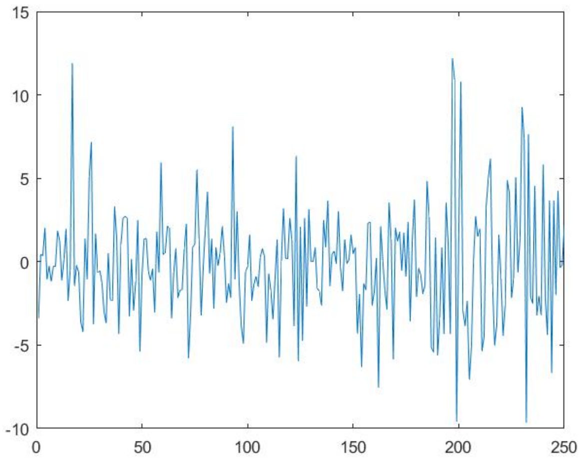

Figure 2: 2018 Goldman Sachs Daily Losses

4 Methodology

In this project we are using three methods to compute the Value at risk (VaR) of our dataset.

Definition: Given a loss L and a confidence level $a \in (0,1)$, the $100a\%$ Value at Risk $VaR_a(L)$ is the quintile of the distribution function $F_L(x) = P(L \leq x)$. In other words that is:

$$VaR_a(L) = F_L^{-1}(a),$$

as we are going to work with continuous distributions and CDFs on this project. One can interpret Value at Risk as the smallest number c : the probability of the loss L exceeding c, is no larger than $1 - a$, where again, $a \in (0, 1)$. [3]

As you understand, the main goal is to fit our data in some distribution for each method, obtain the Maximum Likelihood Estimates of the parameters and then use the inverse of the CDF to obtain the x-value i.e: Value at Risk.

4.1 The Normal Distribution Method

In this method, we are assuming normality, in other words we consider that the losses are normally distributed. We have the parameters of μ as the mean (expectation) of the distribution and σ as the standard deviation.

To obtain Value at Risk with this method, we should first obtain the parameters of the Normal Distribution in which we want to fit our losses.

After putting the data in MATLAB, it was very easy to verify that 2008 was not a good year for the firm, as it has a positive mean on the daily losses and it seems to be very risky too, with high standard deviation. This was due to the economic crisis and the Lehman Brother bankruptcy.

To obtain the quantile of the normal distribution we used the command 'norminv' with confidence level $a = 0.95$ and the mean and standard deviation of each period. That helps us to compute the Value at Risk of the losses.

A small table to summarize our results from this method:

2008	$\mu = 0.0148$	$\sigma = 5.6225$	$VaR_{95} = 9.2629$
2018	$\mu = -0.1150$	$\sigma = 3.4187$	$VaR_{95} = 5.5084$

Table 1: Mean, Standard Deviation and 95% Value at Risk for each period

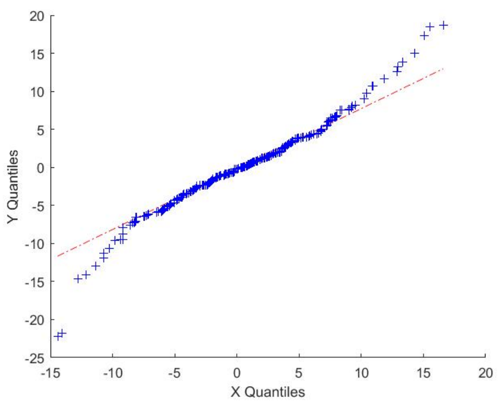

Figure 3: QQ plot of 2008 Losses.

As we expected, Value at Risk is bigger for period 1 (2008). This indicates, that Goldman Sachs expects with 95% probability that the loss will not be more than $9.26. From another viewpoint, Goldman could assume that in approximately 95 out of 100 days of regular trading, the loss will not be larger than the amount in $ stated above. Comparing to 2018, things seem to be better, as the economy 'recovered', so did most of the big Investment Firms. With negative mean on the losses and smaller standard deviation on this period, expected loss for Goldman is of course, lower.

On the other hand, assuming normality for financial data is not always an efficient way to derive results. The reason is simple. From the QQ plot on Figure 3 for 2008 data, one can easily observe that our daily losses are not actually normally distributed.

The QQ-plot has been produced by the inbuilt MATLAB function 'qqplot'. It displays the quantiles from the sample, versus the quantiles from a normal distribution. If the distribution is normal then the plot will appear to be linear. However, we can clearly see that the plot is non-linear which indicates that the losses are not normally distributed.

Having that in mind, we doubt if our results from this Method approximate the reality. The tails of the distribution appear to be different (heavier). That makes us more curious about the results of the next methods, where other distributions are being used.

4.2 The Block Maxima Method

In this method, the idea is to divide the data (losses) into the equal sized blocks and then looking at the largest values inside these blocks.

In our case, we chose as stated above, 250 daily losses for both periods, which then divided in blocks of 5 days each (one can interpret them as a trading week). That gave us 50 blocks, where we are searching for the maximum loss in each of them. After finding those maximum values for each block via some loops, we put them in a vector of 50 values. Then, our assumption for this method is that these losses follow the Generalized Extreme Value distribution:

$$G(x) = e^{-(1+\gamma\frac{x-\mu_n}{\sigma_n})^{-\frac{1}{\gamma}}},$$

Our first goal, before computing Value at Risk, is to estimate the parameters of this distribution: μ_n, σ_n and γ i.e the location, scale and shape parameters. This is the same as estimating:

$$P(M_{50} \leq x)$$

Following our assumption, we fit the Maximum Values of each period in a Generalized Extreme Value Distribution. MATLAB then, gives us the Maximum Likelihood Estimates of the distribution for our fitted data:

2008	$\hat{\mu} = 3.9281$	$\hat{\sigma} = 3.1977$	$\hat{\gamma} = 0.0720$
2018	$\hat{\mu} = 2.4115$	$\hat{\sigma} = 1.8427$	$\hat{\gamma} = 0.1652$

Table 2: Estimates of GEV Parameters for our fitted Max Losses

2008	$VaR_{95} = 14.5181$
2018	$VaR_{95} = 9.4767$

Table 3: 95% Value at Risk using BM for both periods

Generalized Extreme Value distribution has heavier (fatter) tails than Normal distribution. Compared to fat-tailed distributions like GEV, in the normal distribution events that deviate from the mean by five or more standard deviations, have lower probability, meaning that in the normal distribution extreme events are less likely than for fat-tailed distributions. That is why we observe bigger 'Value at Risk' values for the two periods by this method.

One can think of it as an area we try to obtain. When we set 95% confidence, it means that we want our CDF to cover an area of 0.95. In heavy-tailed distributions, that is obtained further on the x-axis than the normal one, making VaR bigger as we saw.

4.3 The Peaks over Thresholds Method

As there exist different models to estimate extreme values, the Peaks over Thresholds (PoT) is one of the models that is frequently used. The distribution of data that is used in PoT method is the Generalized Pareto Distribution. The CDF of this distribution is:

$$H(x) = 1 - (1 + \frac{\gamma}{\sigma}x)^{-1/\gamma}$$

with γ and σ, being the tail index (shape) and the scale parameter, respectively. In the case of $\gamma = 0$ another version of this CDF is used but we do not need it on this project.

The idea on this method is to decide on a Threshold u and look only on the data above this boarder line. We then assume that these points follow the Generalized Pareto Distribution and fit them to obtain the parameter estimates.

We tried to obtain some 'fair' u_1, u_2 thresholds for each period, by shorting our losses from smaller to larger and let only the 5% be above them. Unfortunately, we had some problems with coding this, as MATLAB was in need for more data than the 5%, so we let 20% of the losses to exceed the thresholds for each period. We could have chosen a random (logically random) value for u_i's, but our Max Likelihood Estimates are obtained with the way stated above:

2008	$\hat{\gamma} = 0.0086$	$\hat{\sigma} = 3.7443$	$VaR_{95} = 12.3622$
2018	$\hat{\gamma} = 0.0255$	$\hat{\sigma} = 2.5112$	$VaR_{95} = 8.8179$

Table 4: Estimates of GP Parameters for Losses over u and corresponding Var values

Expected Shortfall

According to definition, Expected Shortfall is the average of all losses which are greater or equal to VaR. A frequent example that is used is the average loss in the $(1-p)\%$ cases, where p is the confidence level. An important feature of ES and VaR that make us to consider both with the same properties, is that both fall when μ increases. The expected shortfall is also known as Conditional Value at Risk.

In this study we go through the estimates from the Peaks over Thresholds method and according to an overview of the Expected Shortfall, it is the expected loss of the days that there exists a Value at Risk (VaR) failure. As there exists different ES and VaR estimates on different estimating methods, it is much better to estimate them together and we must know that the estimate on Es depends on the estimate on VaR. [1]

5 Results

According to the analysis, we can say that assuming normality is not an appropriate model since there is the lack of taking the bad time (fat-tail) periods that encountered like a financial crisis. What we can say about the Block Maxima method is that it 'over-estimates' the losses in compare with the POT method that can take losses better.

6 Appendix

6.1 Daily VaR from Block Maxima VaR

Suppose that we have divided our data in n-days Maximum Losses which follow a GEV distribution with location parameter μ, scale parameter σ and shape parameter γ. Suppose also we have an extremal index θ, which is a measure of dependence of stochastic process values. In this section we are going to see how one can compute the daily $100a\%$ Value at Risk, $a \in (0,1)$ when the distribution of the n-day Maximum Losses is known.

Denote one day loss as L. Our goal, is to compute the daily VaR when the distribution of the n-day Maximum Losses $M_n = max(L_1, ..., L_n)$ is known. Mathematically we want to compute x such that:

$$F_L(x) = P(L \leq x) = a,$$

which means we choose the area i.e the integral of the distribution to be a. Basically, what we are given is:

$$P(M_n \leq x) = P(L \leq x)^{n\theta} = a^{n\theta}$$

But we know that the n-day Maximum Losses have a specific GEV distribution, i.e:

$$P(M_n \leq x) = e^{-(1+\gamma\frac{x-\mu}{\sigma})^{-\frac{1}{\gamma}}}$$

,with given parameters. Therefore:

$$a^{n\theta} = e^{-(1+\gamma\frac{x-\mu}{\sigma})^{-\frac{1}{\gamma}}}$$

Solving for x we obtain:

$$x = \mu - \frac{\sigma}{\gamma}(1 + ln(a^{n\theta})^{-\gamma}),$$

,where μ, σ, γ and probably θ are known, $a \in (0,1)$ is our choice and n depends on our blocks. This x value is the one day $100a\%$ Value at Risk from the Block Maxima.

References

[1] "Overview of Expected Shortfall Backtesting", available online:
$https : //www.mathworks.com/help/risk/overview - of - expected - shortfall - backtesting.html$

[2] Wikipedia, "Normal distribution", available online:
$https : //en.wikipedia.org/wiki/Normal_distribution$

[3] A. Herbertsson, "Static credit portfolio models", Lecture notes (2018) , Page 17-19.