

Back-testing of Expected Shortfall: Main challenges and methodologies



By Leonard BRIE with Benoit GENEST and Matthieu ARSAC

Global Research & Analytics¹

¹ This work was supported by the Global Research & Analytics Dept. of Chappuis Halder & Co.
Many collaborators from Chappuis Halder & Co. have been involved in the writing and the reflection around this paper; hence we would like to send out special thanks to Claire Poinson, Mahdi Kallel, Mikael Benizri and Julien Desnoyers-Chehade

Executive Summary

In a context of an ever-changing regulatory environment over the last years, Banks have witnessed the draft and publication of several regulatory guidelines and requirements in order to frame and structure their internal Risk Management.

Among these guidelines, one has been specifically designed for the risk measurement of market activities. In January 2016, the Basel Committee on Banking Supervision (BCBS) published the Fundamental Review of the Trading Book (FRTB). Amid the multiple evolutions discussed in this paper, the BCBS presents the technical context in which the potential loss estimation has changed from a Value-at-Risk (VaR) computation to an Expected Shortfall (ES) evaluation.

The many advantages of an ES measure are not to be demonstrated, however this measure is also known for its major drawback: its difficulty to be back-tested. Therefore, after recalling the context around the VaR and ES models, this white paper will review ES back-testing findings and insights along many methodologies; these have either been drawn from the latest publications or have been developed by the Global Research & Analytics (GRA) team of Chappuis Halder & Co.

As a conclusion, it has been observed that the existing methods rely on strong assumptions and that they may lead to inconsistent results. The developed methodologies proposed in this paper also show that even though the $ES_{97.5\%}$ metric is close to a $VaR_{99.9\%}$ metric, it is not as easily back-tested as a VaR metric; this is mostly due to the non-elicitability of the ES measure.

Keywords: Value-at-Risk, Expected Shortfall, Back-testing, Basel III, FRTB, Risk Management

EL Classification: C02, C63, G01, G21, G17

Table of Contents

Table of Contents	3
1. Introduction	4
2. Context	4
2.1. Value-at-Risk	5
2.1.1. VaR Definition	5
2.1.2. Risk Measure Regulation	6
2.1.3. VaR Calculation	6
2.1.4. VaR Back-Testing	7
2.2. Expected Shortfall	9
2.2.1. ES Definition	9
2.2.2. ES Regulatory framework	10
2.2.3. ES Calculation	10
2.2.4. VaR vs. ES	11
2.2.5. ES Back-Testing	12
3. ES Back-Testing	14
3.1. Existing Methods	14
3.1.1. Wong's Saddle point technique	14
3.1.2. Righi and Ceretta	17
3.1.3. Emmer, Kratz and Tasche	22
3.1.4. Summary of the three methods	23
3.2. Alternative Methods	25
3.2.1. ES Benchmarking	25
3.2.2. Bootstrap	26
3.2.3. Quantile Approaches	27
4. Applications of the ES methodology and back-testing	31
4.1. ES simulations	31
4.2. Back-test of the ES using our alternative methods	34
5. Conclusion	41

1. Introduction

Following recent financial crises and their disastrous impacts on the industry, regulators are proposing tighter monitoring on banks so that they can survive in extreme market conditions. More recently, the Basel Committee on Banking Supervision (BCBS) announced a change in the Market Risk measure used for Capital requirements in its Fundamental Review of the Trading Book (FRTB), moving from the Value-at-Risk (VaR) to the Expected Shortfall (ES).

However, if the ES captures risks more efficiently than the VaR, it also has one main downside which is its difficulty to be back-tested. This leads to a situation where banks use the ES to perform Capital calculations and then perform the back-testing on a VaR. The focus for banks' research is now to try to find ways to back-test using the ES, as it can be expected that regulators will require so in a near-future.

This paper aims at presenting the latest developments in the field of ES back-testing methodologies and introducing new methodologies developed by the Global Research & Analytics (GRA) team of Chappuis Halder & Co.

First, a presentation of the context in which the back-testing of Expected Shortfall takes place will be provided. This context starts with calculation and back-testing methodologies of the Value-at-Risk, followed by a focus on the ES, analysing its calculation and how it differs from the previous risk measure. The main issues of ES back-testing will then be exposed and discussed.

Second, back-testing methodologies for ES will be reviewed in detail, beginning with methodologies that have already been presented in previous years and then with alternative ones introduced by the research department of Chappuis Halder & Co.

Third, some of the alternative back-testing methodology will be simulated on a hypothetical portfolio and a comparison of the methodologies will be conducted.

2. Context

Recall that in January 2016, the Basel Committee on Banking Supervision (BCBS) issued its final guidelines on the Fundamental Review of the Trading Book (FRTB). The purpose of the FRTB is to cover shortcomings that both regulations and internal risk processes failed to capture during the 2008 crisis. It shows a strategic reversal and the acceptance of regulators for:

- a convergence between risk measurement methods;
- an integrated assessment of risk types (from a silo risk assessment to a more comprehensive risk identification);
- an alignment between prudential and accounting rules.

One of the main requirements and evolutions of the FRTB is the switch from a Value-at-Risk (VaR) to an Expected Shortfall risk measurement approach. Hence, Banks now face the paradox of using the ES for the computation of their Market Risk Capital requirements and the Value-at-Risk for the back-testing. This situation is mainly due to the difficulty of finding an ES back-testing methodology that is both mathematically consistent and practically implementable. However, it can be expected that upcoming regulations will require banks to back-test the ES.

The following sections aim at reminding the existing Market Risk framework for back-testing, as most of the presented notions must be understood for the following chapters of this article. The VaR will therefore be presented at first, given that its calculation and back-testing lay the foundation of this paper. Then, a focus will be made on the ES, by analysing its calculation and the way it differs from the VaR. Finally, the main issues concerning the back-testing of this new measure will be explained. Possible solutions will be the subject of the next chapter.

2.1. Value-at-Risk

2.1.1. VaR Definition

The VaR was first devised by Dennis Weatherstone, former CEO of J.P. Morgan, on the aftermath of the 1987 stock market crash. This new measure soon became an industry standard and was eventually added to Basel I Accord in 1996.

“The Value-at-Risk (VaR) defines a probabilistic method of measuring the potential loss in portfolio value over a given time period and for a given distribution of historical returns. The VaR is expressed in dollars or percentage losses of a portfolio (asset) value that will be equalled or exceeded only X percent of the time. In other words, there is an X percent probability that the loss in portfolio value will be equal to or greater than the VaR measure.

For instance, assume a risk manager performing the daily 5% VaR as \$10,000. The VaR (5%) of \$10,000 indicates that there is a 5% of chance that on any given day, the portfolio will experience a loss of \$10,000 or more.”¹

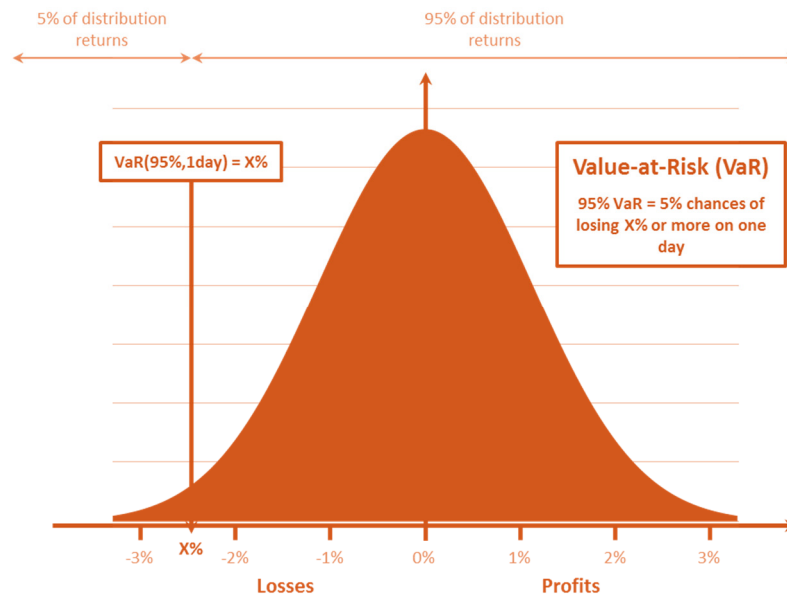


Figure 1 - Probability distribution of a Value-at-Risk with 95% Confidence Level and 1day Time Horizon (Parametric VaR expressed as with a Normal Law $N(0,1)$)

Estimating the VaR requires the following parameters:

¹ Financial Risk Management book 1, Foundations of Risk Management; Quantitative Analysis, page 23

- **The distribution of P&L** – can be obtained either from a parametric assumption or from non-parametric methodologies using historical values or Monte Carlo simulations;
- **The Confidence Level** – the probability that the loss will not be equal or superior to the VaR;
- **The Time Horizon** – the given time period on which the probability is true.

One can note that the VaR can either be expressed in value (\$, £, €, etc.) or in return (%) of an asset value.

The regulator demands a time horizon of 10 days for the VaR. However, this 10 days VaR is estimated from a 1-day result, since a N days VaR is usually assumed equal to the square root of N multiplied by the 1-day VaR, under the commonly used assumption of independent and identically distributed P&L returns.

$$VaR_{\alpha, Ndays} = \sqrt{N} \times VaR_{\alpha, 1day}$$

2.1.2. Risk Measure Regulation

From a regulatory point of view, Basel III Accords require not only the use of the traditional VaR, but also of 3 other additional measures:

- Stressed VaR calculation;
- A new Incremental Risk Charge (IRC) which aims to cover the Credit Migration Risk (i.e. the loss that could come from an external / internal ratings downgrade or upgrade);
- A Comprehensive Risk Measure for credit correlation (CRM) which estimates the price risk of covered credit correlation positions within the trading book.

The Basel Committee has fixed parameters for each of these risk measures, which are presented in the following table:

	VaR	Stressed VaR	IRC	CRM
Confidence Level	99%	99%	99.9%	99.9%
Time Horizon	10 days	10 days	1 year	1 year
Frequency of calculation	Daily	Weekly	-	-
Historical Data	1 previous year	1 stressed year	-	-
Back-Test	Yes	No	-	-

2.1.3. VaR Calculation

VaR calculation is based on the estimation of the P&L distribution. Three methods are used by financial institutions for VaR calculation: one parametric (Variance-Covariance), and two not parametric (Historical and Monte-Carlo).

1. **Variance-Covariance**: this parametric approach consists in assuming the normality of the returns. Correlations between risk factors are constant and the delta (or price sensitivity to changes in a risk factor) of each portfolio constituent is constant. Using the correlation method, the Standard Deviation (volatility) of each risk factor is

extracted from the historical observation period. The potential effect of each component of the portfolio on the overall portfolio value is then worked out from the component's delta (with respect to a particular risk factor) and that risk factor's volatility.

2. **Historical VaR**: this method is the most frequently used method in banks. It consists in applying historical shocks on risk factors to yield a P&L distribution for each scenario and then compute the percentile.
3. **Monte-Carlo VaR**: this approach consists in assessing the P&L distribution based on a large number of simulations of risk factors. The risks factors are calibrated using historical data. Each simulation will be different but in total the simulations will aggregate to the chosen statistical parameters.

For more details about these three methods, one can refer to the Chappuis Halder & Co.'s white paper¹ on the Value-at-Risk.

Other methods such as “Exponentially Weighted Moving Average” (EWMA), “Autoregressive Conditional Heteroskedasticity” (ARCH) or the “declined (G)ARCH (1,1) model” exist but are not addressed in this paper.

2.1.4. VaR Back-Testing

As mentioned earlier, financial institutions are required to use specific risk measures for Capital requirements. However, they must also ensure that the models used to calculate these risk measures are accurate. These tests, also called back-testing, are therefore as important as the value of the risk measure itself. From a regulatory point of view, the back-testing of the risk measure used for Capital requirements is an obligation for banks.

However, in the case of the ES for which no sound back-testing methods have yet been found, regulators had to find a temporary solution. All this lead to the paradoxical situation where the ES is used for Capital requirements calculations whereas the back-testing is still being performed on the VaR. In its Fundamental Review of the Trading Book (FRTB), the Basel Committee includes the results of VaR back-testing in the Capital calculations as a multiplier.

Financial institutions are required to back-test their VaR at least once a year, and on a period of 1 year. The VaR back-testing methodologies used by banks mostly fall into 3 categories of tests: coverage tests (required by regulators), distribution tests, and independence tests (optional).

Coverage tests: these tests assess if the number of exceedances during the tested year is consistent with the quantile of loss the VaR is supposed to reflect.

Before going into details, it seems important to explain how this number of exceedances is computed. In fact, each day of the tested year, the return of the day [t] is compared with the calculated VaR of the previous day [t-1]. It is considered an exceedance if the t return is a loss

¹ Value-at-Risk: *Estimation methodology and best practices*.

greater than the t-1 VaR. At the end of the year, the total number of exceedances during the year can be obtained by summing up all exceedances occurrences.

The main coverage tests are Kupiec’s “proportion of failures” (PoF)¹ and The Basel Committee’s Traffic Light coverage test. Only the latter will be detailed here.

The Traffic Light coverage test dates back to 1996 when the Basel Committee first introduced it. It defines “light zones” (green, yellow and red) depending on the number of exceedances observed for a certain VaR level of confidence. The colour of the zone determines the amount of additional capital charges needed (from green to red being the most punitive).

Zone	Exceptions (out of 250)	Cumulative probability
Green	0	8.11%
	1	28.58%
	2	54.32%
	3	75.81%
	4	89.22%
Yellow	5	95.88%
	6	98.63%
	7	99.60%
	8	99.89%
	9	99.97%
Red	10	99.99%

Table 1 - Traffic Light coverage test (Basel Committee, 1996), with a coverage of 99%

Ex: let’s say a bank chooses to back-test its 99% VaR using the last 252 days of data. It observes 6 exceedances during the year. The VaR measures therefore falls into the “yellow zone”. The back-test is not rejected but the bank needs to add a certain amount of capital.

Distribution tests: these tests (Kolmogorov-Smirnov test, Kuiper’s test, Shapiro-Wilk test, etc.) look for the consistency of VaR measures through the entire loss distribution. It assesses the quality of the P&L distribution that the VaR measure characterizes.

Ex: instead of only applying a simple coverage test on a 99% quantile of loss, we apply the same coverage test on different quantiles of loss (98%, 95%, 90%, 80%, etc.)

Independence tests: these tests assess some form of independence in a Value-at-Risk measure’s performance from one period to the next. A failed independence test will raise doubts on a coverage or distribution back-test results obtained for that VaR measure.

¹ Kupiec (1995) introduced a variation on the binomial test called the proportion of failures (PoF) test. The PoF test works with the binomial distribution approach. In addition, it uses a likelihood ratio to test whether the probability of exceptions is synchronized with the probability “p” implied by the VaR confidence level. If the data suggests that the probability of exceptions is different than p, the VaR model is rejected.

To conclude, in this section were presented the different methodologies used for VaR calculation and back-testing. However, this risk measure has been widely criticized during the past years. Among the different arguments, one can notice its inability to predict or cover the losses during a stressed period, the 2008 crisis unfortunately revealing this lack of efficiency. Also, its incapacity to predict the tail loss (i.e. extreme and rare losses) makes it difficult for banks to predict the severity of the loss encountered. The BCBS therefore decided to retire the well-established measure and replace it by the Expected Shortfall. The following section will aim at describing this new measure and explain how it differs from the VaR.

2.2. Expected Shortfall

The Expected Shortfall (ES), aka Conditional VaR (CVaR), was first introduced in 2001 as a more coherent method than the VaR. The following years saw many debates comparing the VaR and the ES but it's not until 2013 that the BCBS decided to shift and adopt ES as the new risk measure.

In this section are presented the different methodologies of ES calibration and the main differences between the ES and the VaR. Finally, an introduction of the main issues concerning the ES back-testing will be made, which will be the focus of the following chapter.

2.2.1. ES Definition

FRTB defines the ES as the “expected value of those losses beyond a given confidence level”, over certain time horizon. In other words, the t-ES gives the average loss that can be expected in t-days when the returns are above the t-VaR.

For example, let's assume a Risk Manager uses the historical VaR and ES. The observed 97.5% VaR is \$1,000 and there were 3 exceedances (\$1,200; \$1,100; \$1,600). The calibrated ES is therefore \$1,300.

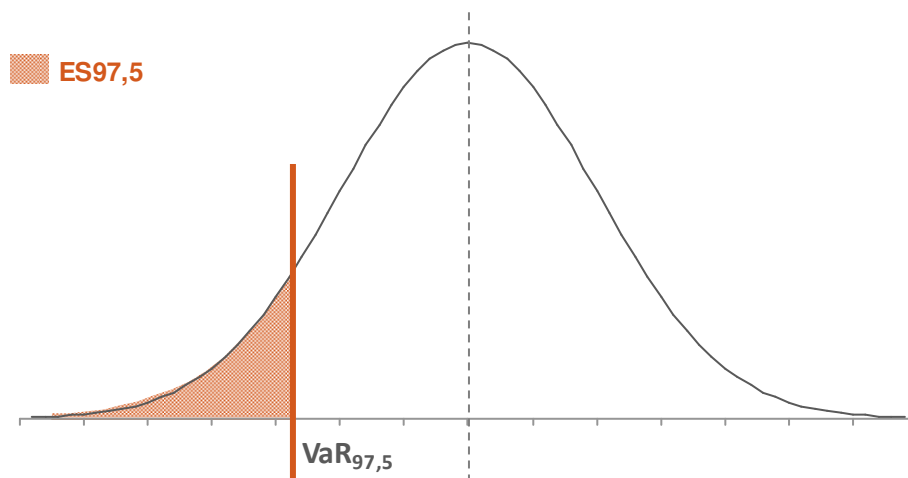


Figure 2 – Expected shortfall (97.5%) illustration

2.2.2. ES Regulatory framework

The Basel 3 accords introduced the ES as the new measure of risk for capital requirement. As for the VaR, the parameters for ES calculation are fixed by the regulators. The following table highlights the regulatory requirements for the ES compared with those of the VaR.

	VaR	Expected Shortfall
Confidence Level	99%	97.5%
Time Horizon	10 days	10 days
Frequency of calculation	Daily	Daily
Historical Data	1 previous year	1 stressed year
Back-Test	Yes	Not for the moment

One can notice that the confidence level is lower for the ES than for the VaR. This difference is due to the fact that the ES is systematically greater than the VaR and keeping a 99% confidence level would have been overly conservative, leading to a much larger capital reserve for banks.

2.2.3. ES Calculation

The calibration of ES is based on the same methodologies as the VaR's. It mainly consists in estimating the right P&L distribution, which can be done using one of the 3 following methods: variance-covariance, historical and Monte-Carlo simulations. These methodologies are described in part 2.1.2.

Once the P&L distribution is known, the Expected Shortfall is calculated as the mean of returns exceeding the VaR.

$$ES_{\alpha,t}(X) = -\frac{1}{1-\alpha} \int_{\alpha}^1 P_t^{-1}(u) du$$

Where :

- X is the P&L distribution;
- t is the time point;
- α is the confidence level;
- $P_t^{-1}(\alpha)$ is the inverse of the VaR function of X at a time t and for a given α confidence level.

One must note that the ES is calibrated on a stressed period as it is actually a stressed ES in the FRTB. The chosen period corresponds to the worst 250 days for the bank's current portfolio in recent memory.

2.2.4. VaR vs. ES

This section aims at showing the main differences (advantages and drawbacks) between the VaR and the ES. The following list is not exhaustive and will be summarized in Table 2:

- **Amount:** given a confidence level $X\%$, the VaR $X\%$ is always inferior to the ES $X\%$, due to the definition of ES as the mean of losses beyond the VaR. This is, in fact, the reason why the regulatory confidence level changed from 99% (VaR) to 97.5% (ES), as banks couldn't have coped with such a high amount of capital otherwise.
- **Tail loss information:** as mentioned earlier, one of the main drawbacks of the VaR is its inability to predict tail losses. Indeed, the VaR predicts the probability of an event but does not consider its severity. For example, a 99% VaR of 1 million predicts that during the following 100 days, 1 loss will exceed 100k, but it doesn't make any difference between a loss of 1.1 million or 1 billion. The ES on the other hand is more reliable as it does give information on the average amount of the loss than can be expected.
- **Consistency:** the ES can be shown as a coherent risk measure contrary to the VaR. In fact, the VaR lacks a mathematical property called sub-additivity, meaning the sum of risk measures (RM) of 2 separate portfolios A and B should be equal or greater than the risk measure of the merger of these 2 portfolios.

$$RM_A + RM_B \geq RM_{A+B}$$

However, in the case of the VaR, one can notice that it does not always satisfy this property which means that in some cases, it does not reflect the risk reduction from diversification effects. Nonetheless, apart from theoretical cases, the lack of sub-additivity of the VaR rarely seems to have practical consequences.

- **Stability:** the ES appears to be less stable than the VaR when it comes to the distribution. For fat-tailed distributions for example, the errors in estimating an ES are much greater than those of a VaR. Reducing the estimation error is possible but requires increasing the sample size of the simulation. For the same error, an ES is costlier than the VaR under a fat-tailed distribution.
- **Cost / Time consumption:** ES calibration seems to require more time and data storage than the VaR's. First, processing the ES systematically requires more work than processing the VaR (VaR calibration being a requirement for ES calculation). Second, the calibration of the ES requires more scenarios than for the VaR, which means either more data storage or more simulations, both of which are costly and time consuming. Third, most banks don't want to lose their VaR framework, having spent vast amount of time and money on its development and implementation. These banks are likely to calculate the ES as a mean of several VaR, which will heavily weigh down calibration time.
- **Facility to back-test:** one of the major issue for ES is its difficulty to be back-tested. Research and financial institutions have been exploring this subject for some years now but still struggle to find a solution that is both mathematically consistent and practically implementable. This difficulty is mainly due to the fact that ES is characterised as "model dependant" (contrary to the VaR which is not). This point is to be explained in the following section.

- Elicibility:** The elicibility corresponds to the definition of a statistical measure that allows to compare simulated estimates with observed data. The main purpose of this measure is to assess the relevance and accuracy of the model used for simulation. To achieve this, one will introduce a scoring function $S(x, y)$ which is used to evaluate the performance of x (forecasts) given some values on y (observations). Examples of scoring functions are squared errors where $S(x, y) = (x-y)^2$ and absolute errors where $S(x, y) = |x - y|$. Given this definition and due to the nature of the Expected Shortfall, one will understand that the ES is not elicitable since there is no concrete observed data to be compared to the forecasts.

	VaR	ES
Amount	✓ -	✓ Originally greater than the VaR, but change of regulatory confidence level from 99% to 97.5%
Tail loss information	✗ Does not give information on the severity of the loss	✓ Gives the average amount of the loss that can be expected
Consistency	✗ Lack of sub-additivity: $VaR_{1+2} > VaR_1 + VaR_2$	✓ Consistent
Stability	✓ Relatively stable	✗ Less stable: the estimation error can be high for some distribution
Cost / Time consumption	✓ -	✗ Always greater than the VaR's
Facility to back-test	✓ Easy to back-test	✗ Difficult to back-test due mainly due to the fact that the back-testing of ES is model dependant
Elicibility	✓ Is elicitable	✗ Isn't elicitable

Table 2 - VaR versus ES: main advantages and drawbacks

2.2.5. ES Back-Testing

As mentioned above, the main issue with the ES is its difficulty to be back-tested. Although research and institutions have already been searching for solutions to this issue for more than 10 years now, no solution seems to satisfy both mathematical properties and practical requirements. Moreover, following the FRTB evolution and its change from VaR to ES for Capital requirements, it has become a priority to be consistent in the use of risk measure (i.e. using the same measure for both Capital calculations and back-testing).

One can wonder why the ES is so difficult to back-test. The main issue is due to the fact that ES back-testing is characterized as model dependent, unlike the VaR which is model independent. Both notions will be described in the following paragraphs.

Let's consider the traditional VaR back-testing and why it is not applicable to the ES. When back-testing VaR, one would look each day at the return "t" to see if it exceeded the $VaR[t-1]$.

The number of exceedances, corresponding to the sum of exceedance occurrences, would then be compared to the quantile the VaR is supposed to reflect.

If one considers that the P&L distribution is likely to change over time, the VaR levels, to which the returns are compared with, also change over time. One would therefore look at the number of exceedances over a value that possibly changes every day. To illustrate this point, the exact same return could be considered as an exceedance one day, and not on another day.

When back-testing the VaR, although the reference value (i.e. the VaR) changes over time, calculating the total number of exceedances still makes sense as one can find convergence of the results. This mathematical property characterizes the VaR back-testing as **model independent**: results are still consistent when the P&L distribution changes during the year.

In the case of the ES however, one would not only look at the number of exceedances but also their values. This additional information complicates the task as there is no convergence when looking at the mean of the exceedances. To make sense, the P&L distribution (or more exactly the VaR) should remain constant during the time horizon. The back-testing of ES is therefore characterized as **model dependent**.

The characterization of ES back-testing as model dependent is one of the main issue that financial institutions experience. Unlike the VaR, they cannot only compare the theoretical ES with the observed value at the end of the year since in most cases the later value does not make sense.

This constraint, combined with limitations in both data storage and time-implementation, makes it difficult for financial institutions and researchers to find new ways to back-test the ES.

The following section aims at presenting the main results and findings of the last 10 years of research and presents alternative solutions introduced by the Global Research & Analytics team of Chappuis Halder & Co.

3. ES Back-Testing

As mentioned earlier, the purpose of this chapter is to present the latest developments in terms of ES back-testing methodologies and to introduce new methodologies developed by the Global Research & Analytics (GRA) team of Chappuis Halder & Co.

3.1. Existing Methods

3.1.1. Wong's Saddle point technique

Wong (2008) proposed a parametric method for the back-testing of the ES. The purpose of the methodology is to find the probability density function of the Expected Shortfall, defined as a mean of returns exceeding the VaR. Once such distribution is found, one can find the confidence level using the Lugannani and Rice formulas, which provide the probability to find a theoretical ES inferior to the sample (i.e. observed) ES. The results of the back-test depend on this “p-value”: given a confidence level of 95%, the p-value must be at least superior to 5% to accept the test.

The method relies on 2 major steps:

1. Inversion formula: find the PDF knowing the moment-generating function
2. Saddle point Technique to approximate the integral

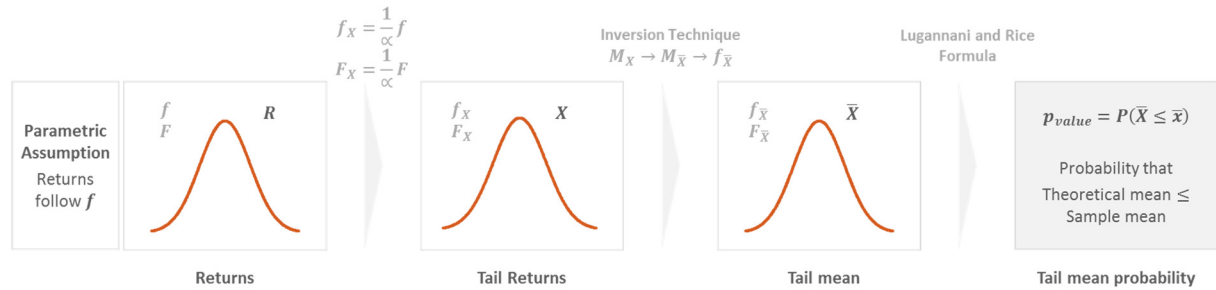


Figure 3 - Overview of Wong's Approach

The ideas of the parametric method proposed by Wong (2008) are as follows. Let $R = \{R_1, R_2, R_3 \dots\}$ be the portfolio returns which has predetermined CDF and PDF denoted by ϕ and f respectively. We denote by $q = \phi^{-1}(\alpha)$ the theoretical α -quantile of the returns.

The statistic used to determine the observed expected shortfall is the following:

$$ES_N^\alpha = -\bar{X} = -\frac{\sum_{t=1}^N R_t I_{\{R_t < q\}}}{\sum_{t=1}^N I_{\{R_t < q\}}}$$

Where $I_{\{x < q\}}$ is the logical test whether the value x is less than the $VaR_\alpha = q$

The purpose of this method is to analytically estimate the density of this statistic and then see where is positioned the observed value with respect to this density.

Are denoted by n , the realised quantity $\sum_{t=1}^N I_{\{R_t < q\}}$ which is the number of exceedances observed in our sample, and X_t the realised return exceedances below the α -quantile q . The observed expected shortfall is then:

$$\widetilde{ES}_N^\alpha = -\bar{x} = -\frac{\sum_{t=1}^n X_t}{n}$$

Reminder of the moment-generating function:

The moment-generating function (MGF) provides an alternative way for describing a random variable, which completely determines the behaviour and properties of the probability distribution of the random variable X :

$$M_X(t) = \mathbb{E}[e^{tX}]$$

The inversion formula that allows to find the density once we have the MGF is the following:

$$f_X(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iut} M_X(iu) du \quad (1.1)$$

One of the known features of the moment-generating function is the following:

$$M_X(\alpha X + \beta Y) = M_X(\alpha t) \cdot M_Y(\beta t) \quad (1.2)$$

PROPOSITION 1:

Let X be a continuous random variable with a density $\alpha^{-1}f(x)$, $x \in (-\infty, q)$. The moment generating function of X is then given by:

$$M_X(t) = \alpha^{-1} \exp(t^2/2) f(q - t) \quad (1.3)$$

and its derivatives with respect to t are given by:

$$M_X'(t) = t \cdot M_X(t) - \alpha^{-1} \cdot \exp(qt) \cdot f(q) \quad (1.4)$$

$$M_X''(t) = t \cdot M_X'(t) + M_X(t) - q \alpha^{-1} \cdot \exp(qt) \cdot f(q) \quad (1.5)$$

$$M_X^{(m)}(t) = t \cdot M_X^{(m-1)}(t) + (m-1) M_X^{(m-1)}(t) - q^{m-1} \alpha^{-1} \cdot \exp(qt) \cdot f(q) \quad (1.6)$$

where $m \geq 3$

Using these, we can also show that the mean and variance of X can be obtained easily:

$$\mu_X = \mathbb{E}[X] = -\frac{f(q)}{\alpha}$$

$$\sigma_X^2 = \text{var}[X] = 1 - \frac{qf(q)}{\alpha} - \mu_X^2$$

The Lugannani and Rice formula

Lugannani and Rice (1980) provide a method which is used to determine the cumulative density function of the statistic \bar{X} (1.1).

It is supposed that the moment-generating function of the variable $X_t = R_t I_{\{R_t < q\}}$ is known. Using the property (1.2), one can compute $M_{\bar{X}}(t) = \left(M_X\left(\frac{t}{n}\right)\right)^n$ and via the inversion formula will obtain:

$$f_{\bar{X}}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \left(M_X\left(i\frac{t}{n}\right)\right)^n dt = \frac{n}{2\pi} \int_{-\infty}^{\infty} e^{n(K[it]-itx)} dt$$

where $f_{\bar{X}}(x)$ denotes the PDF of the sample mean and $K[t] = \ln M_X[t]$ is the cumulative-generating function of $f_X(x)$. Then the tail probability can be written as:

$$P(\bar{X} > \bar{x}) = \int_{\bar{x}}^q f_{\bar{X}}(t) dt = \frac{1}{2\pi i} \int_{\Omega-i\infty}^{\Omega+i\infty} e^{n(K[t]-t\bar{x})} \frac{dt}{t}$$

where Ω is a saddle-point¹ satisfying:

$$K'(\Omega) = \bar{x} \quad (1.9)$$

The saddle-point is obtained by solving the following expression deduced from (1.4) and (1.5):

$$K'(\Omega) = \frac{M'(\Omega)}{M(\Omega)} = \Omega - \exp(q\Omega - \Omega^2/2) \frac{f(q)}{\phi(q - \Omega)} = \bar{x} \quad (1.10)$$

Finally, Lugannani and Rice propose to approximate this integral as follows:

PROPOSITION 2:

Let Ω be a saddle-point satisfying the equation (1.9) and define:

$$\epsilon = \Omega \sqrt{nK''(\Omega)}$$

$$\delta = \text{sgn}(\Omega) \sqrt{2n \left(\Omega \widetilde{E S}_N^\alpha - K(\Omega) \right)}$$

where $\text{sgn}(\Omega)$ equals to zero when $\Omega = 0$, or takes the value of $1/(-1)$ when $\Omega < 0/(\Omega > 0)$. Then the tail probability of \bar{x} less than or equal to the sample mean \bar{x} is given by

$$P(\bar{X} \leq \bar{x}) = \begin{cases} \phi(\delta) - f(\delta) \cdot \left(\frac{1}{\epsilon} - \frac{1}{\delta} + O(n^{-3/2}) \right) & \text{for } \bar{x} < q \text{ and } \bar{x} \neq \mu_X \\ 1 & \text{for } \bar{x} > q \\ -\frac{1}{2} + \frac{K^{(3)}(0)}{6\sqrt{2\pi n(K''(\Omega))^3}} + O(n^{-3/2}) & \text{for } \bar{x} = \mu_X \end{cases}$$

¹ In mathematics, a saddle point or minimax point is a point on the surface of the graph of a function where the slopes (derivatives) of orthogonal function components defining the surface become zero (a stationary point) but are not a local extremum on both axes.

Once the tail probability is obtained, one can compute the observed expected shortfall \widetilde{ES}_N^α and carry out a one-tailed back-test to check whether this value is too large. The null and alternative hypotheses can be written as:

$$H_0: \widetilde{ES}_N^\alpha = \overline{ES}_N^\alpha \quad \text{versus} \quad H_1: \widetilde{ES}_N^\alpha > \overline{ES}_N^\alpha$$

where \overline{ES}_N^α denotes the theoretical expected shortfall under the null hypothesis.

The p-value of this hypothesis test is simply given by the Lugannani and Rice formula as

$$p_{value} = P(\bar{X} \leq \bar{x})$$

Example:

For a portfolio composed of one S&P500 stock, it is assumed that the bank has predicted that the daily P&L log-returns are i.i.d and follow a normal distribution calibrated on the observations of the year 2014. Then all the observations of the year 2015 are normalised so one can consider that the sample follows a standard normal distribution $\mathcal{N}(0,1)$. Using Wong’s method described above, the steps to follow in order to back-test the ES under these assumptions and with $\alpha = 2.5\%$ are:

1. Calculate the theoretical α -quantile: $q = -\phi^{-1}(2.5\%) = 1.96$
2. Calculate the observed ES of the normalized log-returns of 2015: $\bar{X} = -2.84, n = 19$
3. Solve the equation (1.10) to find the saddle-point: $\Omega = -3.23$
4. Calculate $K[\Omega]$ and $K''[\Omega]$ where

$$K''[t] = \frac{d}{dt} \frac{M'(t)}{M(t)} = \frac{M''(t)M(t) - M'(t)^2}{M(t)^2}$$

In our case, we found: $K[\Omega] = 8.80$ and $K''[\Omega] = 2.49$

5. Calculate the tail probability of \widetilde{ES}_N^α and compare to the level of confidence tolerated by the p_{value} test: $P(ES_N^\alpha \leq \widetilde{ES}_N^\alpha) \sim 0$

In this example, the null hypothesis is rejected. Not only does it show that the movements of 2015 cannot be explained by the movements of 2014, but it also shows that the hypothesis of a normal distribution of the log-returns is not likely to be true.

3.1.2. Righi and Ceretta

The method of Righi and Ceretta is less restrictive than Wong’s one in the sense that the law of the returns may vary from one day to another. However, it requires the knowledge of the truncated distribution below the negative VaR level.

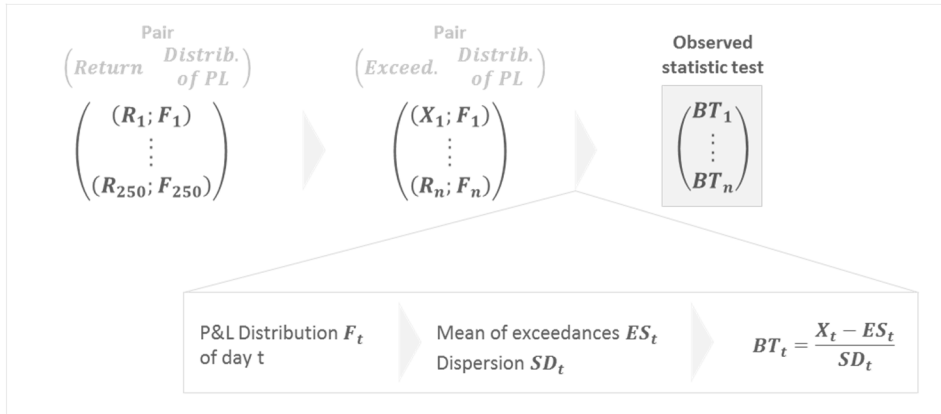


Figure 4 - Righi and Ceretta - Calculating the Observed statistic test

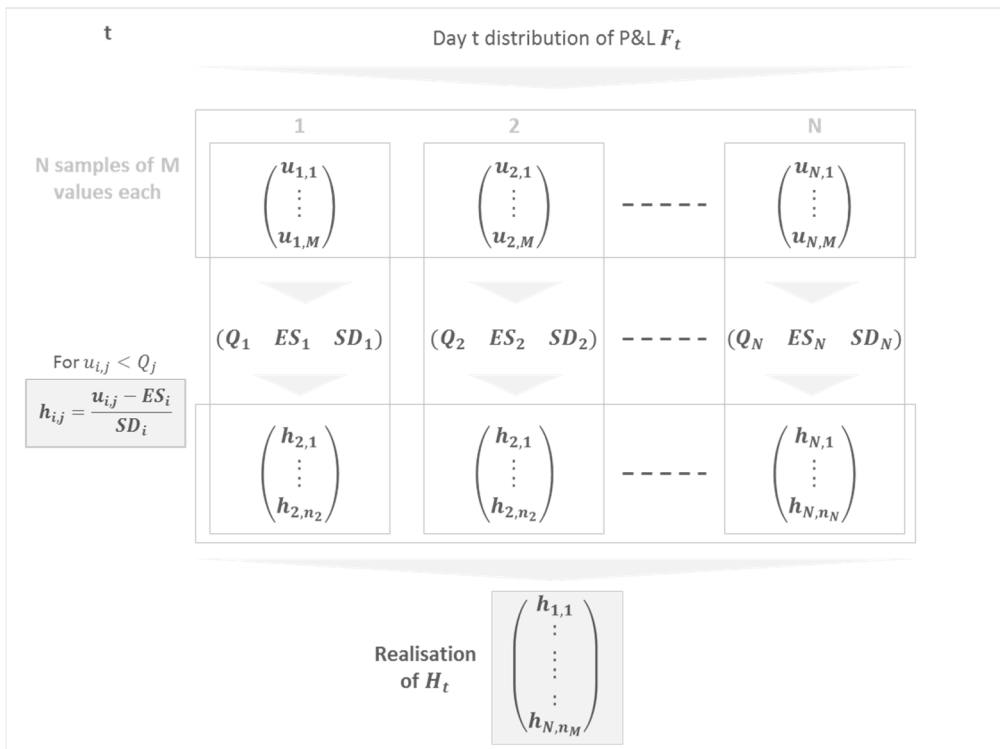


Figure 5 - Righi and Ceretta - Simulating the statistic test

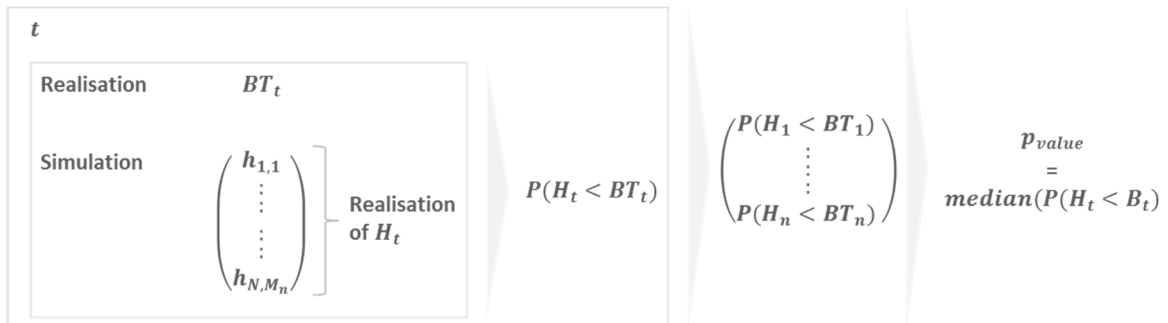


Figure 6 - Righi and Ceretta - Overview

In their article, they consider that the portfolio log-returns follow generalized autoregressive conditional heteroscedastic ($GARCH(p, q)$) model, which are largely applied in finance:

$$r_t = \mu_t + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t$$

$$\sigma_t^2 = \omega + \sum_p \rho_i \varepsilon_{t-i}^2 + \sum_q \beta_j \sigma_{t-j}^2$$

where r_t is the log-return, μ_t is the conditional mean, σ_t is the conditional variance and ε_t is the shock over the expected value of an asset in period t ; ω , ρ and β are parameters; z_t represents the white noise series which can assume many probability distribution and density functions denoted respectively F_t and f_t .

The interest behind using this model is mainly the fact that we can easily predict the truncated distribution properties, mainly the α – quantile/ES:

$$Q_{\alpha,t} = \mu_t + \sigma_t F^{-1}(\alpha)$$

$$ES_t = \mu_t + \sigma_t \mathbb{E}[z_t | z_t < F^{-1}(\alpha)] \quad (2.1)$$

But one can also calculate the dispersion of the truncated distribution as follows:

$$SD_t = \sqrt{\text{Var}(\mu_t + \sigma_t z_t | z_t < F^{-1}(\alpha))} = \sigma_t \sqrt{\text{Var}(z_t | z_t < F^{-1}(\alpha))} \quad (2.2)$$

The ES and SD are mainly calculated via Monte-Carlo simulations. In some cases, it is possible to have their parametric formulas:

1. Case where z_t is normal |

It is assumed that z_t is a standard Gaussian noise $\mathcal{N}(0,1)$, which is a very common case. The expectation of the truncated normal distribution is then:

$$\mathbb{E}[z_t | z_t < Q] = \frac{f(Q)}{F(Q)}$$

substituting this expression in the equation of (2.1), it is obtained:

$$ES_t = \mu_t + \sigma_t \frac{f(F^{-1}(\alpha))}{\alpha}$$

The variance of a truncated normal distribution below a value Q is given by:

$$\text{VAR}[z_t | z_t < Q] = 1 - Q \frac{f(Q)}{F(Q)} - \left(\frac{f(Q)}{F(Q)} \right)^2$$

Substituting this expression in the variance term of the formula (2.2), it is deduced:

$$SD_t = \sigma_t \cdot \left[1 - F^{-1}(\alpha) \frac{f(F^{-1}(\alpha))}{\alpha} - \left(\frac{f(F^{-1}(\alpha))}{\alpha} \right)^2 \right]^{\frac{1}{2}}$$

2. Case where z_t follows a Student's distribution |

It is assumed that z_t is a Student's t distributed random variable with ν degrees of freedom. One can show that the truncated expectation is as follow:

$$\mathbb{E}[z_t | z_t < Q] = \frac{1}{2\sqrt{\nu} F(Q) \beta\left(\frac{\nu}{2}, \frac{1}{2}\right)} \left(Q^2 GH\left(\frac{1+\nu}{2}, 1; 2; -\frac{Q^2}{2}\right) \right)$$

substituting this expression in the expectation of (1), it is obtained:

$$ES_t = \mu_t + \sigma_t \left(\frac{1}{2\sqrt{\nu}\alpha \cdot \beta\left(\frac{\nu}{2}, \frac{1}{2}\right)} \left(F^{-1}(\alpha)^2 GH\left(\frac{1+\nu}{2}, 1; 2; -\frac{F^{-1}(\alpha)^2}{2}\right) \right) \right)$$

where $\beta(\cdot, \cdot)$ and $GH(\cdot, \cdot; \cdot; \cdot)$ are the Beta and Gauss hyper geometric functions conform to:

$$\beta(a, b) = \int_0^1 u^{a-1} (1-u)^{b-1} du$$

$$GH(a, b; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k z^k}{(c)_k k!}$$

Where $(\cdot)_k$ denotes the ascending factorial.

Similarly, for the standard normal SD, it is deduced from the variance of a truncated Student's t distribution:

$$VAR[z_t | z_t < Q] = \frac{1}{3\sqrt{\nu} F(Q) \beta\left(\frac{\nu}{2}, \frac{1}{2}\right)} Q^3 GH\left(\frac{1+\nu}{2}, \frac{3}{2}; \frac{5}{2}; -\frac{Q^2}{2}\right)$$

Again, substituting this variance term in (2.2), one will obtain an analytical form of the standard deviation of the truncated distribution:

$$SD_t = \sigma_t \cdot \left[\frac{1}{3\sqrt{\nu}\alpha \cdot \beta\left(\frac{\nu}{2}, \frac{1}{2}\right)} F^{-1}(\alpha)^3 GH\left(\frac{1+\nu}{2}, \frac{3}{2}; \frac{5}{2}; -\frac{F^{-1}(\alpha)^2}{2}\right) \right]^{\frac{1}{2}}$$

Once the ED and SD are expressed and computed, for each day in the forecast period for which a violation in the predicted Value at Risk occurs, the following test statistic is defined:

$$BT_t = \frac{r_t - ES_t}{SD_t} \quad (2.3)$$

$$HT = \frac{z_t - \mathbb{E}[z_t | z_t < Q]}{\sqrt{VAR[z_t | z_t < Q]}} \quad (2.4)$$

Where z_t is the realisation of the random variable Z_t (in the Garch process, it is supposed that Z_t is iid but it is not necessarily the case).

The idea of Righi and Ceretta is to see where the value of BT_t is situated with respect to the “error” distribution of the estimator $HT = \frac{Z_t - \mathbb{E}[Z_t|Z_t < Q]}{\sqrt{VAR[Z_t|Z_t < Q]}}$ by calculating the probability $\mathbb{P}(HT < BT_t)$ and then take the median (or eventually the average) of these probabilities over the time as a p-value over a certain confidence level p .

They propose to calculate this distribution using Monte-Carlo simulations following this algorithm:

- 1) Generate N times a sample of $n - iid$ random variable u_{ij} under the distribution F , $i = 1, \dots, n$; $j = 1, \dots, N$;
- 2) Estimating for each sample the quantity $\mathbb{E}[u_{ij}|u_{ij} < q(u_{ij})]$ and $VAR[u_{ij}|u_{ij} < q(u_{ij})]$ where $q(u_{ij})$ is the α -th worst observation over the sample u_{ij}
- 3) Calculate for each realisation u_{ij} , the quantity $h_{ij} = \frac{u_{ij} - \mathbb{E}[u_{ij}|u_{ij} < q(u_{ij})]}{\sqrt{VAR[u_{ij}|u_{ij} < q(u_{ij})]}}$ which is a

realisation of the random variable HT defined above.

- 4) Given the actual BT_t , estimate $\mathbb{P}(H_t < BT_t)$ using the sample h_{ij} as an empirical distribution of H_t
- 5) Determine the test p-value as the median of $\mathbb{P}(H_t < BT_t)$ and compare the value to the test level fixed at p .

The methodology has been applied on the test portfolio of the normalized daily returns for the 2014 to 2015 year. The results, where z_t is a standard Gaussian noise $\mathcal{N}(0,1)$, are the following:

Inputs	Distribution*	Standard Normal
	Level of freedom*	none
	Confidence Level of the ES*	97,5%
	Scenario	05/08/2014
BT Results	VaRth	1,96
	ESth	2,34
	VaRobs	2,38
	Number of exceedances	12
	ESobs	2,49
	var(X<-VaRobs)	0,0941
	Critical Value** - median	0,00%
Critical Value** - mean	0,00%	
Output	Final Output	PASS

Table 3 Summary of the Righi and Ceretta implementation

For the test value of a fixed level at 97.5%, the Righi and Ceretta methodology gives satisfactory results with a pass for both the median and mean computations. **Finally, one can conclude that this methodology is acceptable, nevertheless it relies on a parametric assumption that may not fit the portfolio, which is not captured by the test statistics.**

In the table below are displayed the exceedance rates and the associated test statistics:

Exceedance #	Exceedance Value	Test statistic	P(Ht<Bt)
1-	2,128	1,179	2,4%
2-	2,128	1,180	2,9%
3-	2,304	0,606	1,7%
4-	1,996	1,610	4,3%
5-	2,361	0,420	1,4%
6-	2,879	- 1,270	0,3%
7-	2,831	- 1,111	0,4%
8-	3,082	- 1,930	0,2%
9-	3,077	- 1,916	0,2%
10-	2,681	- 0,623	0,6%
11-	2,396	0,306	1,3%
12-	2,014	1,550	4,1%

Table 4 Exceedance rates and test statistics of the portfolio

3.1.3. Emmer, Kratz and Tasche

The method presented by Emmer and al. (2013) consists in replacing the ES back-testing with a VaR back-testing. This substitution relies on the approximation of the ES as a mean of several VaR levels, according to the following formula:

$$ES_{\alpha} = \frac{1}{1-\alpha} \int_{\alpha}^1 VaR_u du = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{k=0}^{N-1} VaR_{\alpha+k \left(\frac{1-\alpha}{N}\right)}$$

$$\approx \frac{1}{5} \left(VaR_{\alpha} + VaR_{\alpha+\frac{1-\alpha}{5}} + VaR_{\alpha+2 \cdot \frac{1-\alpha}{5}} + VaR_{\alpha+3 \cdot \frac{1-\alpha}{5}} + VaR_{\alpha+4 \cdot \frac{1-\alpha}{5}} \right)$$

Hence, assuming $\alpha=97.5\%$, the formula becomes:

$$ES_{97.5\%} = \frac{1}{5} (VaR_{97.5\%} + VaR_{98\%} + VaR_{98.5\%} + VaR_{99\%} + VaR_{99.5\%})$$

Therefore, by back-testing the different VaR 97.5%, 98%, 98.5%, 99% and 99.5% one should complete the back-testing of ES. If all these levels of VaR are validated, then the ES should be considered as well.

However, this methodology has many drawbacks since one should determine an appropriate N that ensures that the average VaR converges to the ES, otherwise it would imply too many uncertainties in the approximation. Given the value of N, the tests could not be implemented due to high computation time (calculation of N different VaRs). For instance, in the Emmer and al. proposal, it is assumed that the convergence is obtained for N=5 which explains the means performed on 5 Value-at-Risk.

Finally, one should also propose an adapted traffic light table since it may be not relevant or too restrictive to require a pass on all the VaR levels.

3.1.4. Summary of the methods

In this section, it is summarised the three different methods in term of application and implementation as well as their drawbacks:

Wong's method I

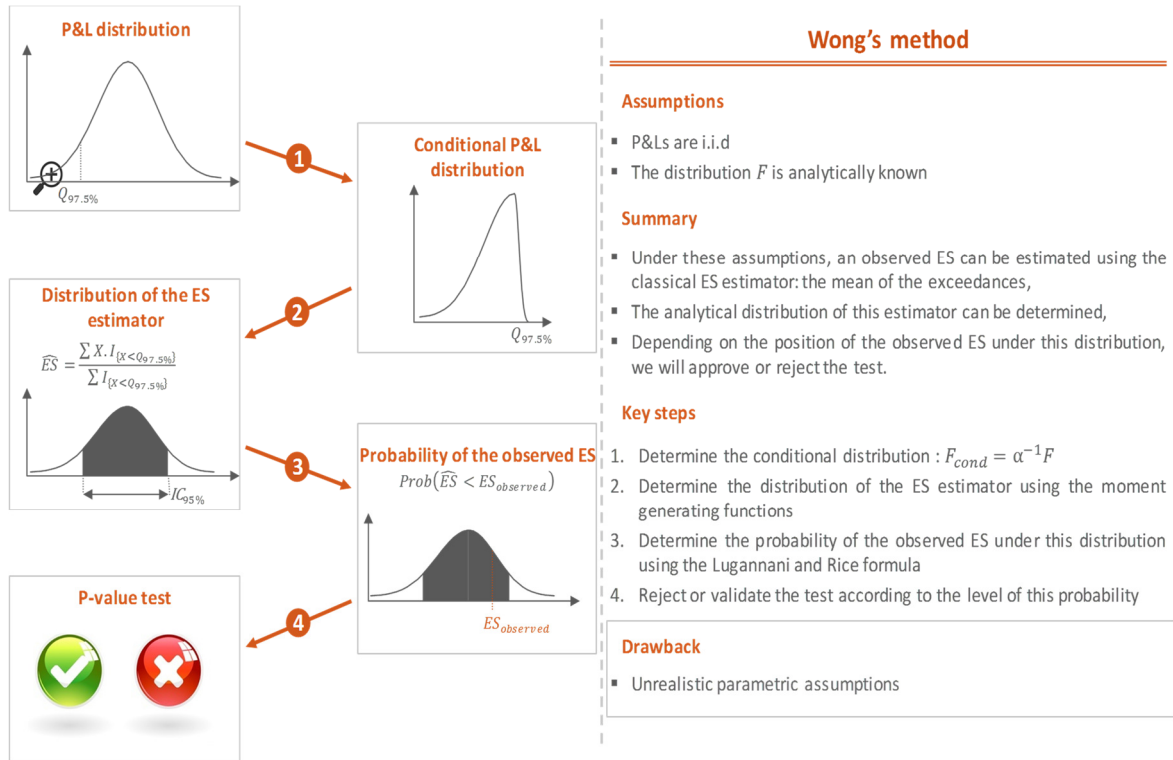


Figure 7 - Summary of Wong's methodology

Righi and Ceretta Method |

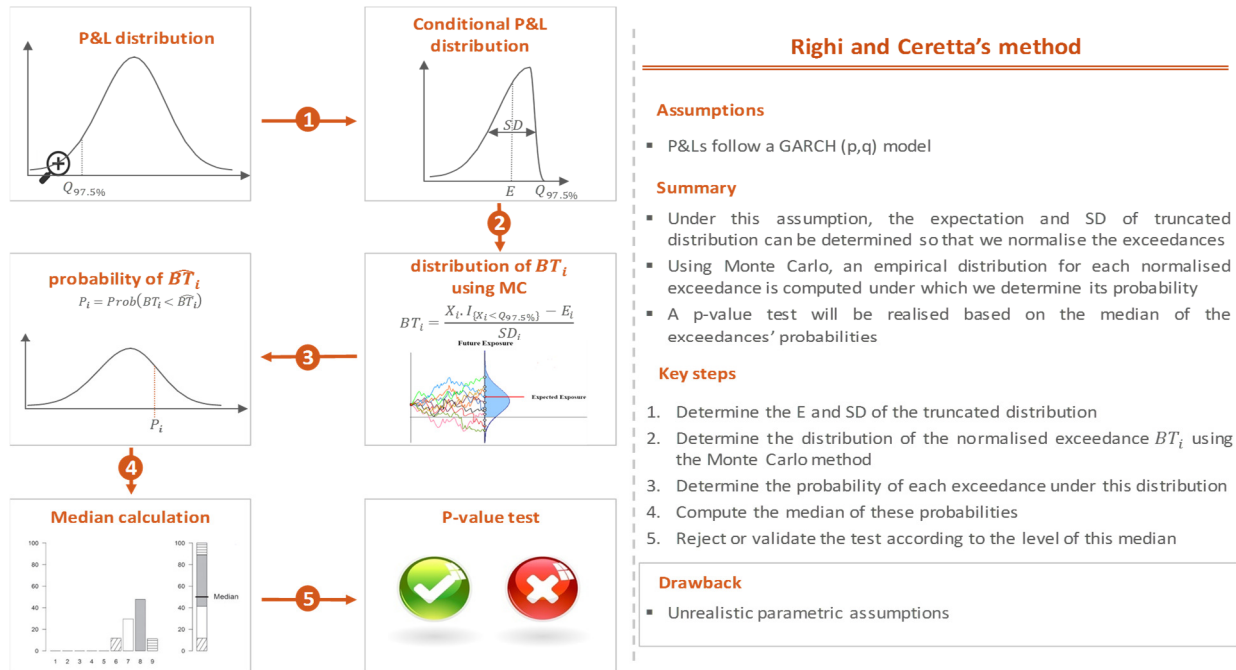


Figure 8 - Summary of Righi and Ceretta's methodology

Emmer, Kratz and Tasche Method |

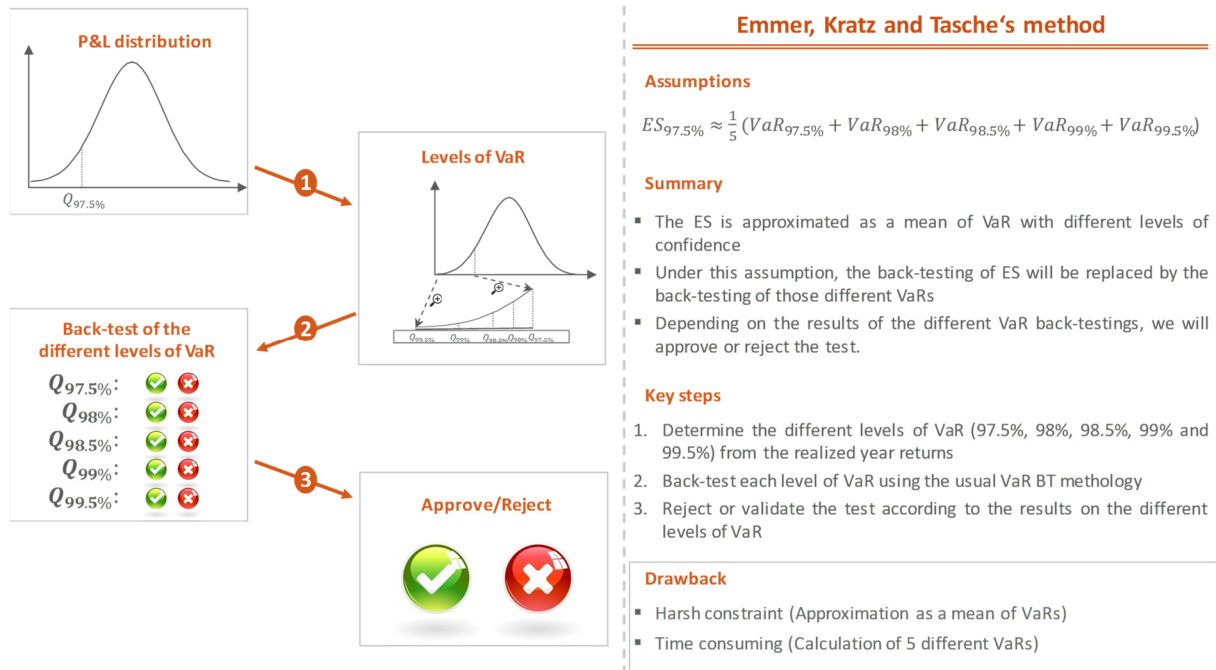


Figure 9 - Summary of Emmer, Kratz and Tasche's methodology

3.2. Alternative Methods

In the following sections are presented alternative methods introduced by the Global Research & Analytics (GRA) department of Chappuis Halder & Co.

First of all, it is important to note that some of the following methods rely on a major hypothesis, which is the consistency of the theoretical VaR for a given period of time. This strong – and not often met - assumption is due to the use of what is called “observed ES”.

The observed ES reflects the realised average loss (above the 97.5% quantile) during a 1-year time period as illustrated in the below formula:

$$ES_{obs} = \frac{\sum_{t=1}^{250} X_{t+1} I(X_{t+1} > VaR_t)}{N}$$

Where X_t corresponds to the return day t and N is the number of exceedances during the year ($N = \sum_{t=1}^{250} I(X_{t+1} > VaR_t)$ with I the identity function)¹.

However, this value only makes sense as long as the theoretical VaR (or more broadly the P&L distribution used for calibration) doesn't change during this time period. Should the opposite occur, one would look at the loss beyond a level that changes with time, and the average of these losses would lose any meaning.

3.2.1. ES Benchmarking

This method focuses on the distance $ES_{th} - ES_{obs}$ between the theoretical ES (obtained by calibration) and the observed ES (corresponding to realised returns). The main goal of this methodology is to make sure the distance $ES_{th} - ES_{obs}$ (back-testing date) is located within the confidence interval. This interval can be found by recreating a distribution from historical values. The output of the back-test depends on the position of the observed distance: if the value is within the interval, the back-test is accepted, otherwise it is rejected.

The historical distribution is based on 5 years returns (i.e. 5*250 values). For each day of these 5 years, the distance $ES_{th} - ES_{obs}$, ES_{th} and ES_{obs} is calculated as described in the introduction of this section. The 1,250 values collected can therefore be used to build a distribution that fits historical behaviour.

¹ As mentioned in part 2.1.1, VaR is calculated with a 1-day time horizon. Therefore, the return that is compared to the VaR[t] is the return X[t+1]

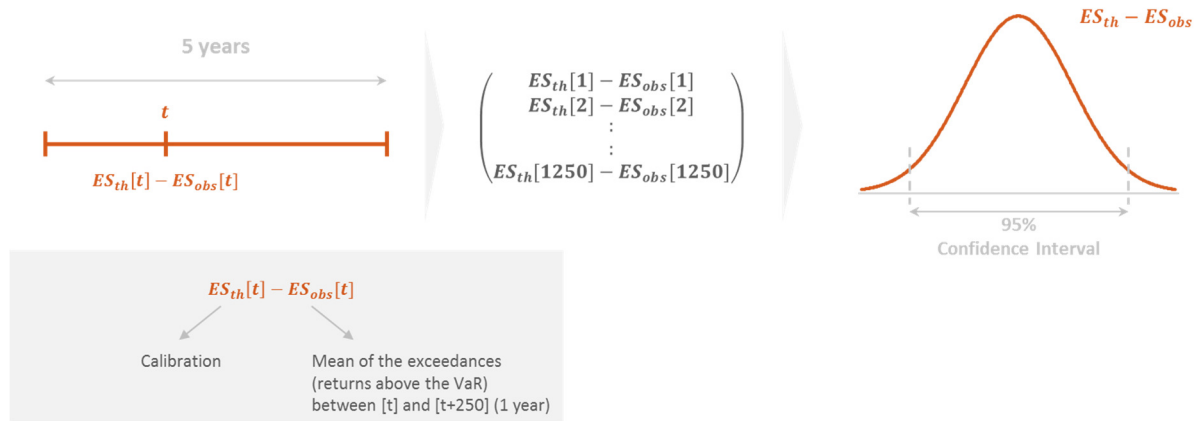


Figure 10 - Illustration of ES Benchmarking Methodology

One of the main downside of the methodology is that it relies on the notion of observed ES. However, as mentioned earlier, this particular value requires a constant VaR, which is not often met in reality.

Finally, once the confidence interval is obtained, one can use it in order to back-test the simulated ES on the future back-testing horizon.

3.2.2. Bootstrap

This methodology focuses on the value of the observed ES. As for the previous methodology, the goal of this method is to verify that the observed ES is located within the confidence interval. The latest can be found by recreating a distribution from historical values using the bootstrap approach which is detailed below. The output of the test depends on the position of the observed ES (back-testing date): if the value is in the interval, the back-test is accepted, otherwise it is rejected.

In this methodology, the bootstrap approach is used to build a more consequent vector of returns in order to find the distribution of the ES as the annual mean of returns exceeding the VaR. This approach consists in simulating returns, using only values from a historical sample. The vector formed by all simulated values therefore only contains historical data that came from the original sample.

The overall methodology relies in 3 steps as illustrated in Figure 11:

1. The **sample vector** is obtained and contains the returns of 1-year data;
2. Use of the bootstrap method to create a bigger vector, filled only with values from the sample vector. This vector will be called the “**Bootstrap Vector**”;
3. The **final vector**, used for compiling the distribution, is obtained by selecting only returns exceeding the VaR from the bootstrap vector;
4. The **distribution** can be reconstructed, using the final vector.

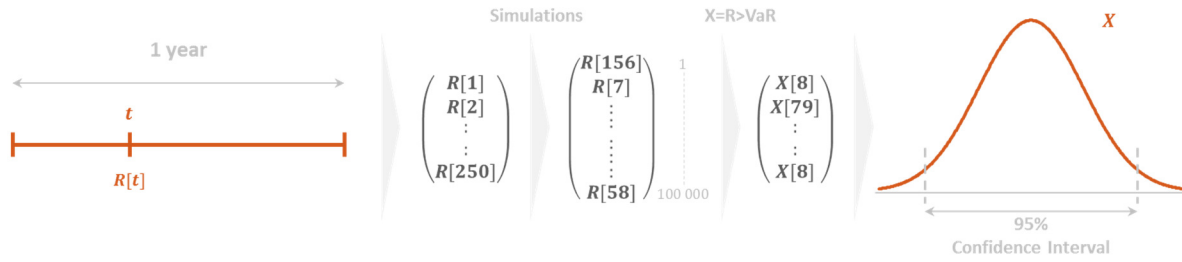


Figure 11 - Illustration of Bootstrap Methodology

3.2.3. Quantile Approaches

Whereas the Expected Shortfall is usually expressed as a value of the loss (i.e. in £, \$, etc.), the two methodologies Quantile 1 and Quantile 2 choose to focus on the ES as a quantile, or a probability value of the P&L distribution. The two methods differ in the choice of the quantile adopted for the approach.

The following paragraphs describe the two different options for the choice of the quantile.

Quantile 1

This methodology focuses on the quantile of the observed ES (back-testing date), in other words the answer to the question: “which probability is associated to the value of a specific Expected Shortfall in the P&L distribution?”

One must notice that this quantile is not the confidence level of the Expected Shortfall. Indeed, let’s take a confidence level of 97.5% as requested by the regulation. It is possible to estimate an observed VaR and therefore an observed ES as a mean of returns exceeding the VaR. The observed quantile can be found by looking at the P&L distribution and spotting the probability associated to the ES value. The ES being strictly greater than the VaR, the quantile will always be strictly greater than 97.5%.

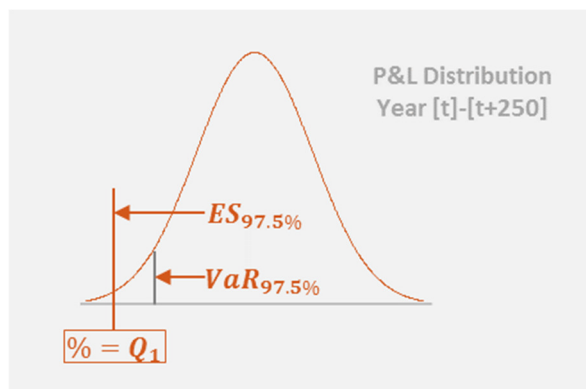


Figure 12 - Calculation of the Quantile Q_1

Quantile 2

This methodology looks at the stressed-retroactive quantile of the observed ES (back-testing date), that is to say the answer to the question: “To which quantile correspond the observed ES

at the time of the back-testing if it was observed in the reference stressed period used for calibration?”



Figure 13 - Calculation of the quantile Q_2

Back-testing methodology

Once the choice of the quantile computation is done, the approach is the same for the two methodologies: it consists in verifying that the calculated quantile at the date of the back-testing is located in the confidence interval obtained from a reconstructed historical distribution. If the quantile is within the confidence interval, the back-test is accepted, otherwise it is rejected.

The distribution is obtained using the same framework as for the ES Benchmarking methodology (see Section 3.2.1). The quantile is computed each day over 5 years (for the first method the observed quantile and for the second one, the stressed-retroactive quantile). Those 1,250 values are used to build a historical distribution of the chosen quantile and the confidence interval immediately follows.

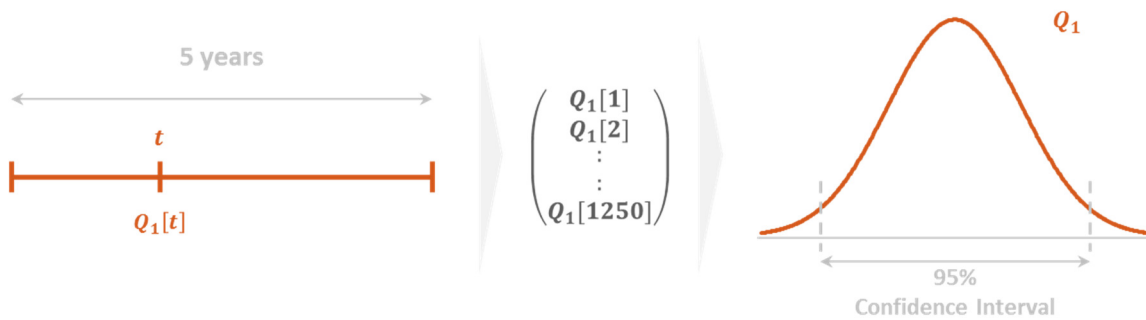


Figure 14 - Illustration of the Quantile 1 Methodology

3.2.4. Summary of the methods

ES Benchmarking |

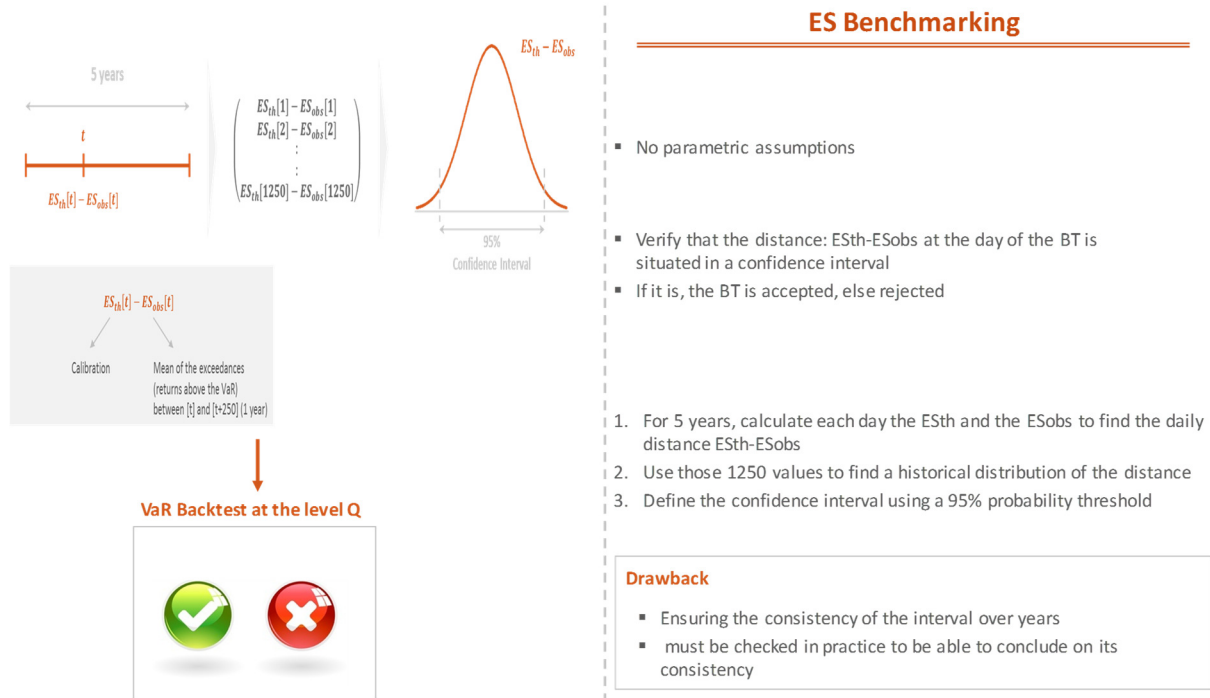


Figure 15- Summary of the ES Benchmarking methodology

Quantile method I

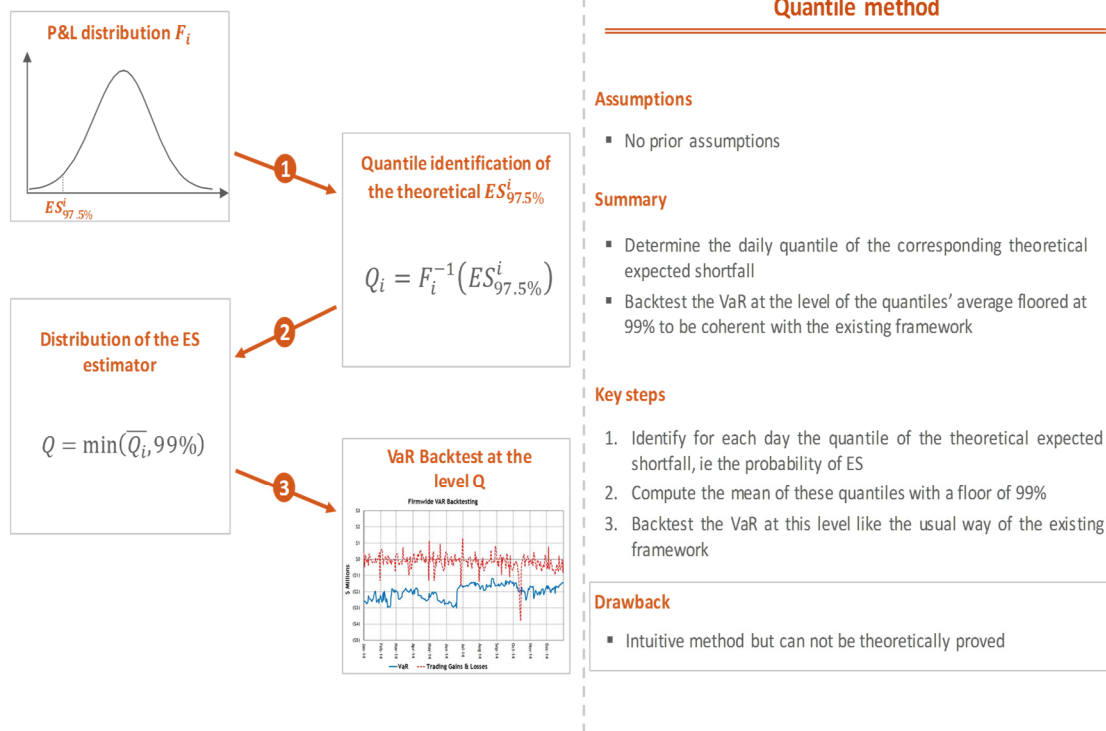


Figure 16 - Summary of the Quantile methodology

Bootstrap method I

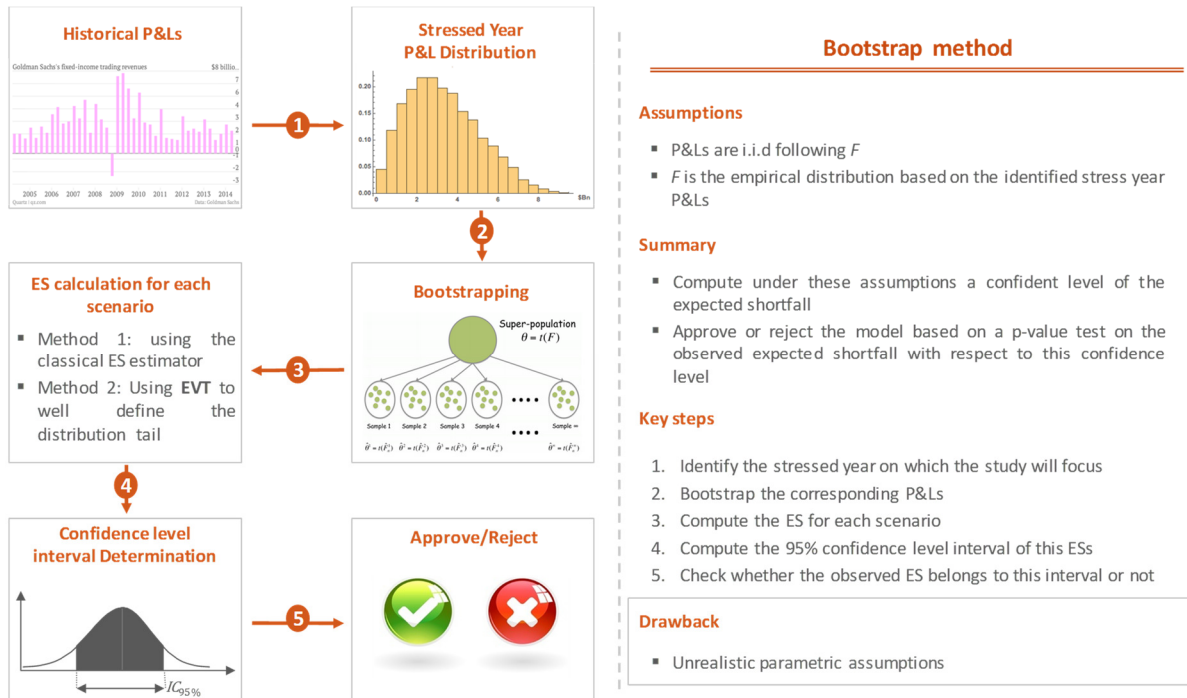


Figure 17 - Summary of the Bootstrap methodology

4. Applications of the ES methodology and back-testing

4.1. ES simulations

In this section, the back-testing approaches presented in the 3.2 section have been applied on simulations of the S&P 500 index¹. Indeed, instead of performing back-testing on parametric distributions; it has been decided to perform a back-testing exercise on simulated values of an equity (S&P 500 index) based on Monte Carlo simulations. The historic levels of the S&P 500 are displayed in Figure 18 below:

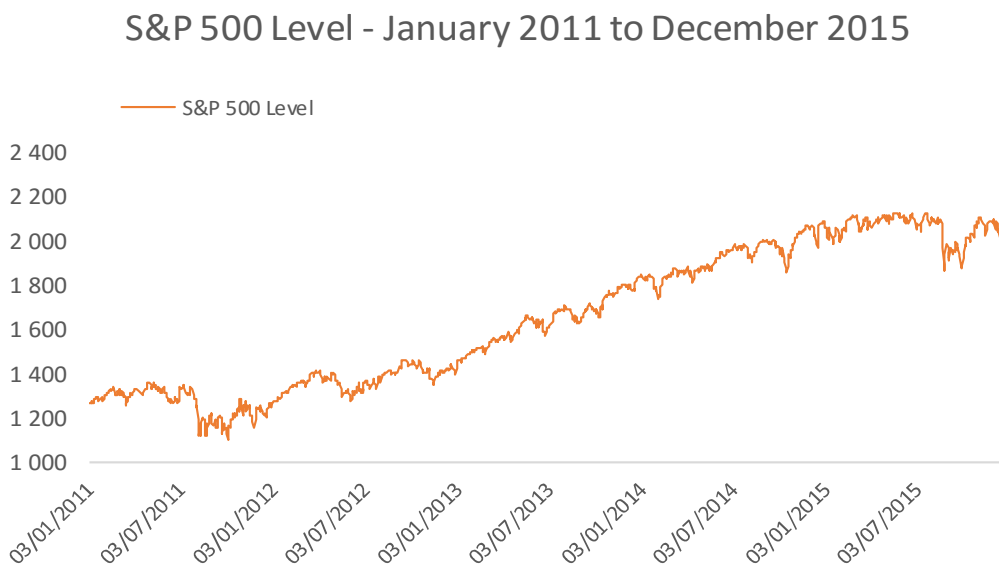


Figure 18 - S&P 500 Level – From January 2011 to December 2015

A stochastic model has been used in order to forecast the one-day return of the stock price which has been compared to the observed returns. **The stochastic model relies on a Geometric Brownian Motion (hereafter GBM) and the simulations are done with a daily reset as it could be done in a context of Market Risk estimation.**

The stochastic differential equation (SDE) of a GBM in order to diffuse the stock price is as follows:

$$dS = S(\mu dt + \sigma dW)$$

And the closed form solution of the SDE is:

$$S(t) = S(0)e^{\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W(t)\right]}$$

Where:

- S is the Stock price

¹ The time period and data selected (from January 2011 to December 2015) is arbitrary and one would obtain similar results and findings with other data.

- μ is the expected return
- σ is the standard deviation of the expected return
- t the time
- $W(t)$ is a Brownian Motion

Simulations are performed on a day-to-day basis over the year 2011 and 1,000 scenarios are produced per time points. **Therefore, thanks to these simulations, it is possible to compute a one-day VaR_{99%} as well as a one-day ES_{97.5%} of the return which are then compared to the observed return price.**

Both VaR and ES are computed as follows:

$$VaR_{99\%}(t) = Q_{99\%} \left(\frac{S_{sim}(t) - S_{obs}(t-1)}{S_{obs}(t-1)} \right)$$

$$ES_{97.5\%}(t) = \frac{\sum_{i=1}^n \left(\frac{S_{sim}(t) - S_{obs}(t-1)}{S_{obs}(t-1)} \right) I_{\left\{ \frac{S_{sim}(t) - S_{obs}(t-1)}{S_{obs}(t-1)} \geq VaR_{97.5\%}(t) \right\}}}{\sum_{i=1}^n I_{\left\{ \frac{S_{sim}(t) - S_{obs}(t-1)}{S_{obs}(t-1)} \geq VaR_{97.5\%}(t) \right\}}}$$

Where n is the total number of scenarios per time point t.

The figure below shows the results of our simulations and computations of the VaR_{99%} and ES_{97.5%}:

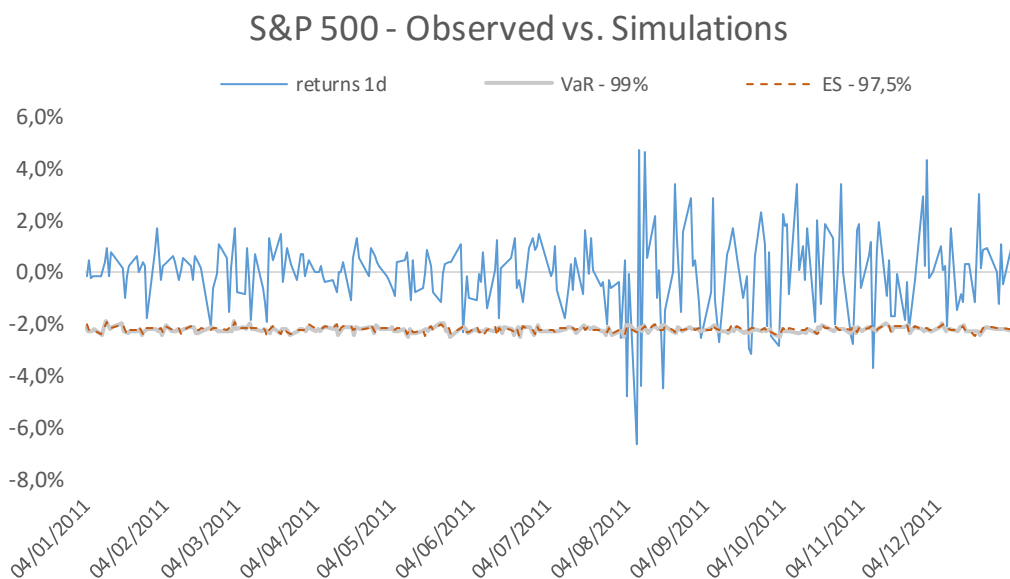


Figure 19 - Observed returns vs. simulations – From January 2011 to January 2012

Figure 19 shows that in comparison to the observed daily returns, the VaR_{99%} and the ES_{97.5%} gives the same level of conservativeness. This is further illustrated with Figure 20 where it is observed that the level of VaR and Expected shortfall are close.

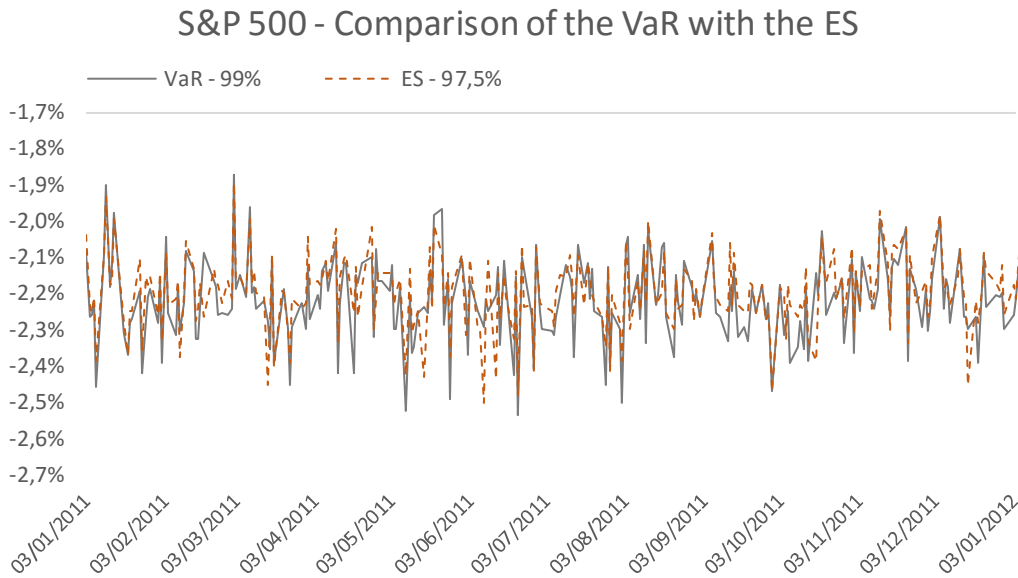


Figure 20 – Comparison of the VaR_{99%} with the ES_{97.5%} - From January 2011 to January 2012

When looking at Figure 20, one can notice that the ES_{97.5%} doesn't always lead to more conservative results in comparison to the VaR_{99%}. This is explained by the fact that the ES is the mean of the values above the VaR_{97.5%}, consequently and depending on the Monte Carlo simulations it is realistic to observe ES_{97.5%} slightly below the VaR_{99%}.

Finally, when looking at the simulations, one can conclude that both risk measures are really close. Indeed the distribution of the spread between the simulated VaR_{99%} and the ES_{97.5%} (see Figure 21 below); it is observed that 95% of the spread between both risk measures are within the]-0.1%, 0.275%] interval.

Distribution of the spread of the (VaR_{99%} - ES_{97,5%}) - From January 2011 to December 2015

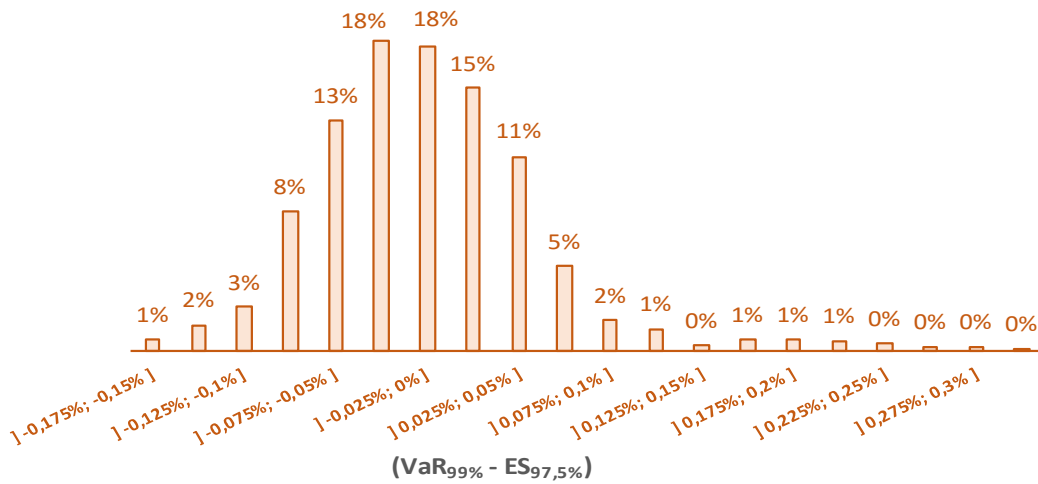


Figure 21 – Spread VaR_{99%} vs. ES_{97,5%} - January 2011 to September 2015

Following the computation of these simulated ES, it can be concluded that in comparison to a VaR measure, the ES is not overly conservative and severe measure. Given these findings and knowing that the ES is a more consistent measure in comparison to the VaR (due to the way it is estimated), it can be accepted as a suitable risk measure provided that a reliable approach is used in order to back-test the results.

4.2. Back-test of the ES using our alternative methods

Following the previous conclusions, it has been decided to focus on some approaches defined in section 3.2. That's why an observed ES has been computed based on the daily VaR_{97,5%} (obtained via the MC simulations) and the observed returns over the year following the simulation date. Its expression is as follows:

$$ES_{obs}(t) = \frac{\sum_{i=1}^m R(t+i) I_{\{R(t+i) \geq VaR_{97,5\%}(t)\}}}{\sum_{i=1}^m I_{\{R(t+i) \geq VaR_{97,5\%}(t)\}}}$$

Where m is the number of days in the year following the date t and R(t) the daily return observed at date t:

$$R(t) = \frac{S_{obs}(t) - S_{obs}(t-1)}{S_{obs}(t-1)}$$

As presented in the section 3.2.1, this observed ES has been compared to the theoretical daily simulated ES.

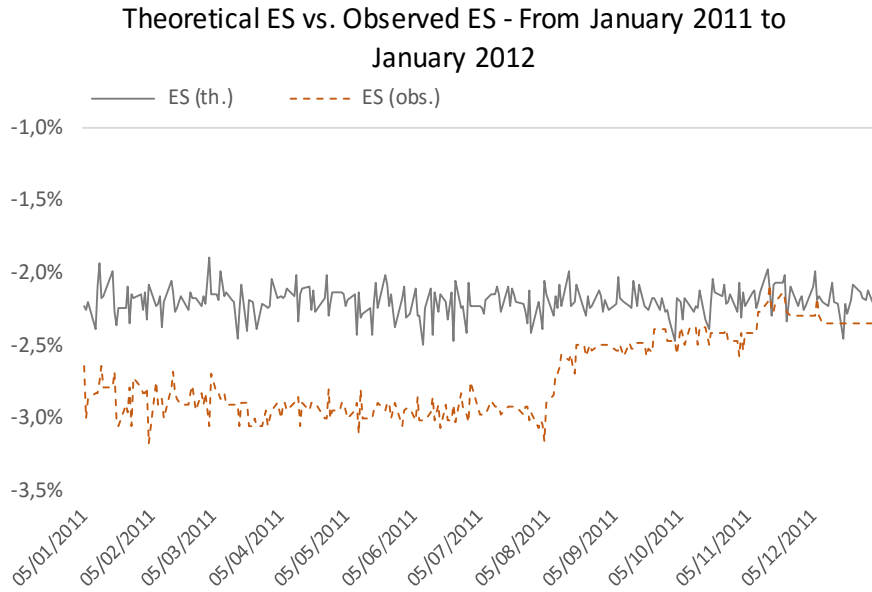


Figure 22 – Comparison of the theoretical ES against the observed ES - From January 2011 to January 2012

Figure 22 shows both theoretical and observed ES computed over the year 2011 whereas Figure 23 presents the distribution of the distance $ES_{th} - ES_{obs}$.

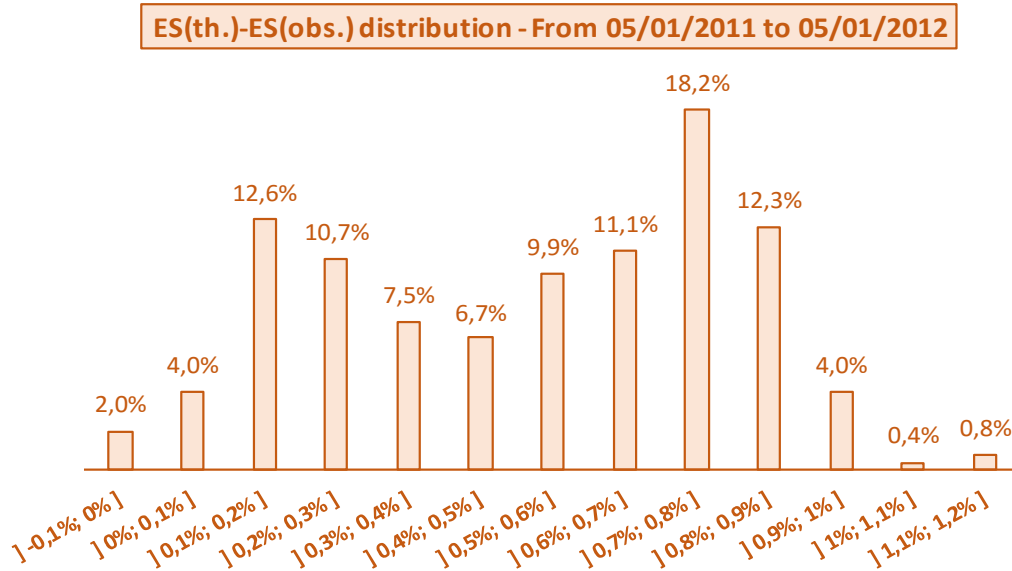


Figure 23 – Distribution of the distance $ES_{th} - ES_{obs}$ - From January 2011 to January 2012

Figure 23 shows that, over the year 2011, the observed ES is lower than the theoretical ES in 98% of the case and 95% of the distance range between [0.04%; 0.95%]. Based on these results, it is possible to define a 95% confidence interval for the future comparison of the observed ES vs. the theoretical ES in order to assess the accuracy and conservativeness of the theoretical ES. This confidence interval has been applied to the data of the year 2012, 2013 and 2014 where:

- a **positive result** is obtained when the distance between the theoretical and observed ES is below the lower bound (i.e. the theoretical ES is conservative),
- a **neutral result** is obtained when the distance is within the confidence interval (the theoretical and observed ES are assumed close),
- a **negative result** is obtained when the distance is above the upper bound (i.e. the theoretical ES lack of conservativeness).

Results are presented in Table 5 below, where it is noted that the interval computed over the year 2011 leads to satisfactory results since a majority of positive results are observed.

Year	Positive		Neutral		Negative	
	#	%	#	%	#	%
2012	72	28,7%	179	71,3%	-	0,0%
2013	174	69,3%	77	30,7%	-	0,0%
2014	159	63,3%	92	36,7%	-	0,0%
Total	405	53,7%	348	46,2%	0	0,0%

Table 5 – Results of the ES back-testing – 2012, 2013 and 2014

The benefits of this approach is that it gives a way to back-test the level of the simulated ES via the computation of thresholds based on the results of the previous year.

Nevertheless, one can challenge the way the observed ES is computed. Indeed, instead of relying on a forward-looking approach; it could be computed via a backward-looking approach:

$$ES_{obs}(t) = \frac{\sum_{i=1}^m R(t-i) I_{\{R(t-i) \geq VaR_{97.5\%}(t)\}}}{\sum_{i=1}^m I_{\{R(t-i) \geq VaR_{97.5\%}(t)\}}}$$

This approach has been tested on the 2013 data. Figure 24 shows both theoretical and observed ES computed over the year 2013 using a backward-looking approach whereas Figure 25 the results of the forward-looking methodology.

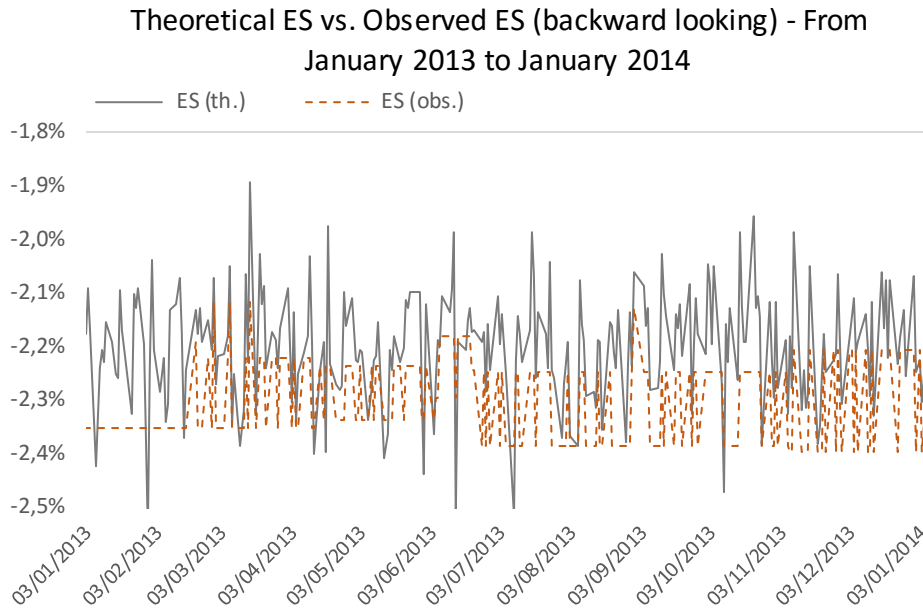


Figure 24 – Comparison of the theoretical ES against the observed ES (backward looking)- January 2013 to January 2014

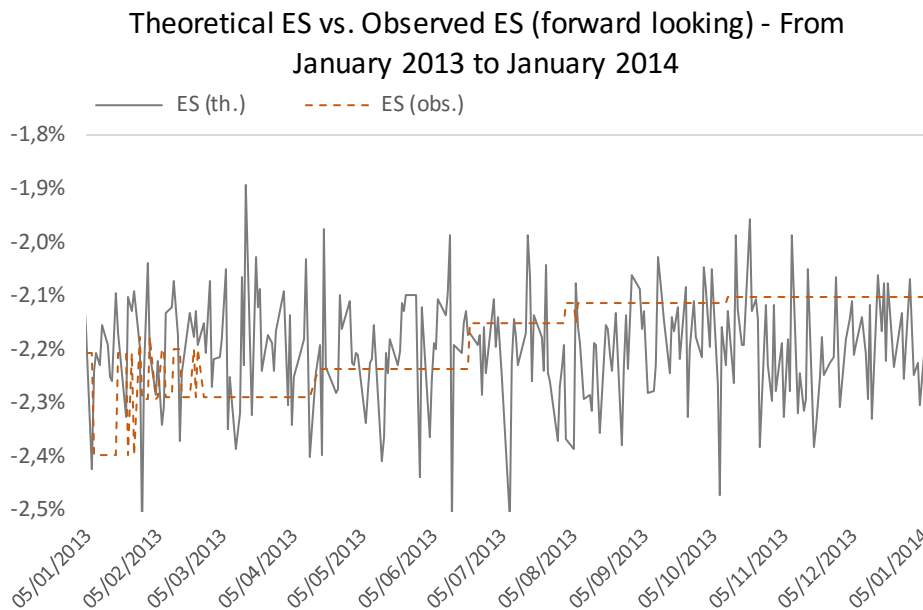


Figure 25 – Comparison of the theoretical ES against the observed ES (forward looking)- January 2013 to January 2014

The comparison of both Figure 24 and Figure 25 reveals that the backward-looking approach leads to a more conservative and consistent computation of the observed ES since the distance between the simulations and the observations is marginal. **Furthermore, the use of a backward-looking approach can be implemented on a daily basis, whereas the forward-looking relies on future observation of returns.**

The results of the backward-looking approach have been used in order to recalibrate the interval, as expected the new interval is now narrowed and is equal to $[-0.07\%; 0.26\%]$. The results of the ES back-testing are presented in Table 6:

Year	Positive		Neutral		Negative	
	#	%	#	%	#	%
2014	78	31,1%	172	68,5%	1	0,4%
2015	141	56,2%	21	8,4%	89	35,5%
Total	227	30,1%	495	49,3%	282	28,1%

Table 6 – Results of the ES back-testing (backward looking) – 2014 and 2015

When looking at Table 6, one can notice that the ES back-testing relying on the backward-looking approach leads to more situations where the simulated ES are underestimated which is explained by the interval being smaller.

Overall, this shows the complexity of back-testing the ES since it is less straight forward than a VaR back-testing and based on the definition of the observed ES. Furthermore, it can be noted in Table 5 and Table 6 that when using the ES as a risk measure it could lead to instability in the back-testing results over years, which shows the importance of defining a proper back-testing methodology.

Then, it has been decided to test the bootstrap alternative method presented in the Section 3.2. As a first step, a **sample vector** corresponding to the one-day returns of the year 2011 has been computed. As a second step, the **bootstrap vector** has been constructed; this vector is filled with the values of the sample vector that have been selected randomly ten thousand times. Figure 26 below shows the **bootstrap vector distribution**:

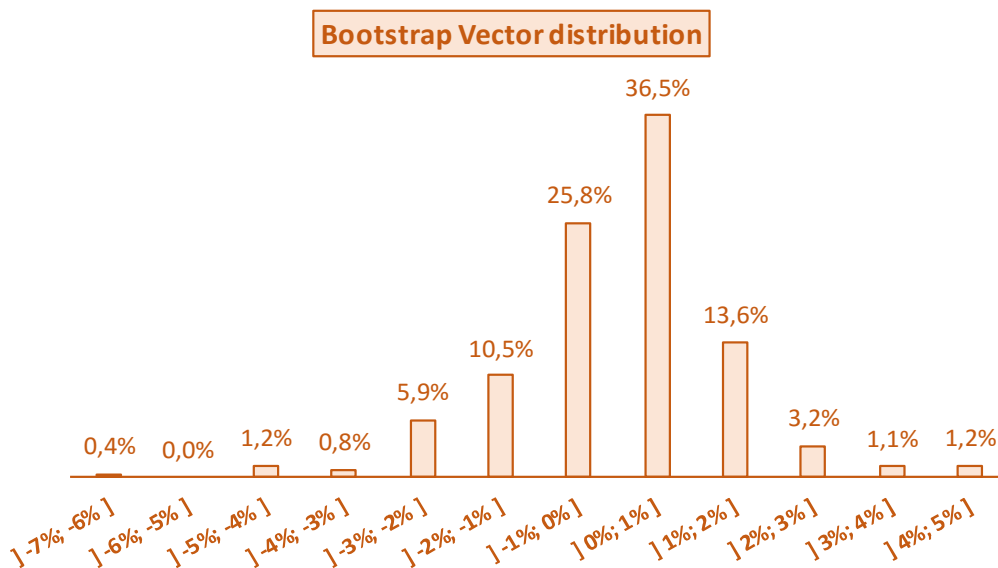


Figure 26 – Bootstrap Vector – Random sampling of the 2011 one-day returns - 10 000 observations

Finally, for each date of the year 2011, a **final vector** with all the value exceeding the daily estimated VaR is estimated. For instance, as of January 6 2011, the estimated $VaR_{97.5\%}$ is -1.97% which leads to a **final vector** distribution as follows:

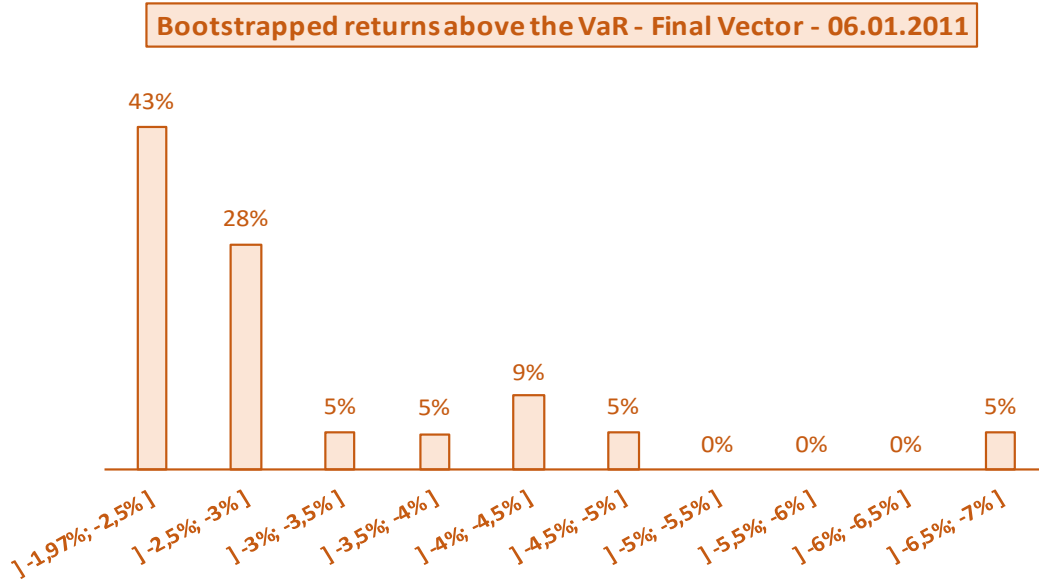


Figure 27 – Final Vector – 06.01.2011 – $VaR_{97.5\%} = -1.97\%$

As such, for each time point of the year 2011, it is possible to estimate a Bootstrapped $ES_{97.5\%}$ which will be used as a reference value to back-test the simulated $ES_{97.5\%}$. The results of the ES back-testing are presented in Figure 28 below:

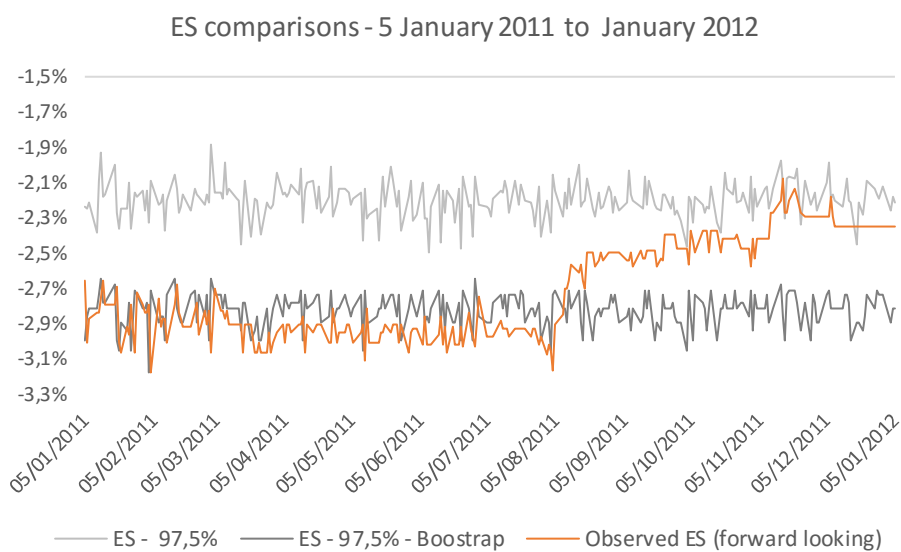


Figure 28 – ES Comparisons – January 2011 to January 2012

When looking at Figure 28, one can compare the bootstrapped $ES_{97.5\%}$ and the Observed ES as reference value for back-testing purpose. In particular, it is observed that both curves are similar for the first 9 months, then the observed ES significantly decreases from this date. **Hence, the bootstrapped ES appears to be more stable over the year and would be a more reliable value to back-test the simulated ES since no breach are observed in comparison to the observed ES where it fails on the last months.** The same exercise has been done on the data of the year 2015 and with an observed ES computed using a backward-looking approach. Results are displayed in Figure 29 where similar conclusions are drawn.

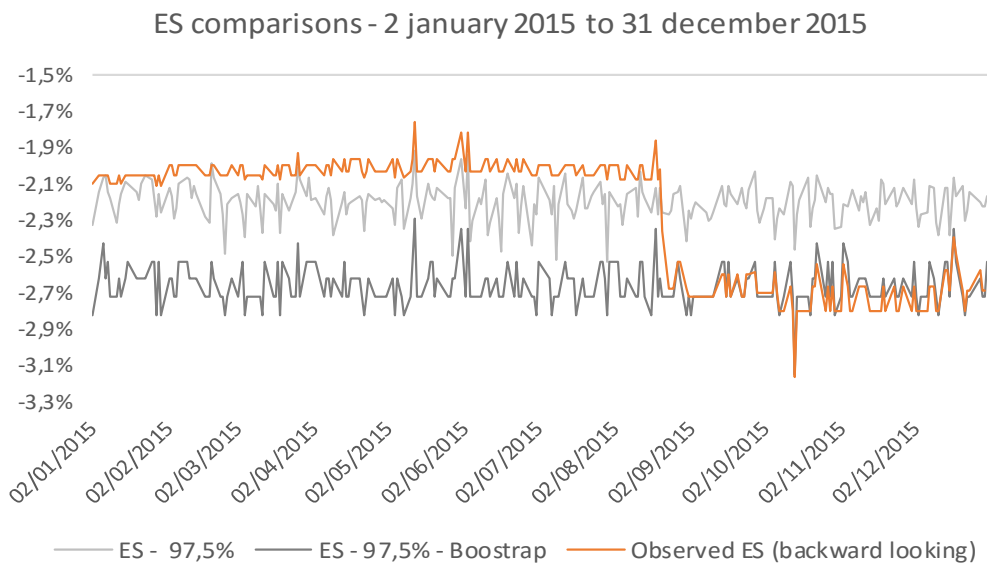


Figure 29 - ES Comparisons – 02 January 2015 to 31 December 2015

5. Conclusion

This white paper presented the latest developments in terms of ES back-testing methodologies and introduced new methodologies developed by the Global Research & Analytics (GRA) team of Chappuis Halder & Co. After presenting and testing several methods that can be found in the literature, it has been concluded that these methods may not fit for the purpose of a regulatory back-testing since they rely on questionable assumptions or heavy computation time.

Then, in order to highlight the specificities of the back-testing of the Expected Shortfall, it has been decided to implement and test the alternative methods that have been presented in this article. Overall, it has been concluded that the complexity of back-testing the Expected Shortfall relies on a proper definition of the observed ES, which should serve as a reference value for back-testing. Indeed, it is clear that the estimation of a simulated Expected Shortfall is quite straightforward since it relies on the computation of the simulated Value-at-Risk; this is not the case of the computation of the observed Expected Shortfall. Indeed, in order to perform an apple-to-apple comparison, one can't just compare a simulated daily Expected Shortfall to a daily observed return. Knowing that the Expected shortfall corresponds to the average value above the worst loss defined under a specific quantile, it sounds natural to introduce these features while estimating the observed Expected Shortfall.

Hence, in order to propose a relevant back-testing of the simulated ES, one should first decide on the assumptions used for the computation of the observed ES. For example, it is important to choose if it has to be computed with a backward or forward-looking approach, the number of time points to use, the frequency of the calculations, etc.

These assumptions need to be chosen wisely in order to calibrate a relevant interval of confidence for ES comparisons. Indeed, it has been discussed in this article that the back-testing results could be different and instable with regards to the computation methodology of the observed ES.

That's why, on the basis of the tests performed in this article, it has been observed that the more reliable back-testing results came from the computation of a bootstrapped ES since it has the advantage of considering a P&L distribution constant during the time horizon, which produced a stable but conservative level of confidence.

References

- *Consultative document Fundamental Review of the trading book: A revised market risk framework*, Basel Committee on Banking Supervision, January 2014
- *Minimum capital requirements for market risk*, Basel Committee on Banking Supervision, January 2016
- *Introducing three model-independent, non-parametric back-test methodologies for Expected Shortfall*, Carlo Acerbi and Balazs Szekely, December 2014
- *Individual and Flexible Expected Shortfall Backtesting*, Marcelo Righi and Paulo Sergio Ceretta, June 2013
- *Backtesting Expected Shortfall: the design and implementation of different backtests*, Lisa Wimmerstedt, August 2015
- *Backtesting trading risk of commercial banks using expected shortfall*, Woon K Wong, 2008
- *Techniques for verifying the accuracy of risk measurement models*, P.H. Kupiec, 1995