alfi association of the luxembourg fund industry

guidelines

# ABC OF VAR MODEL BACKTESTING

## A Practitioners' Guide

in association with ALRiM LUXEMBOURG

# table of contents

# Executive summary

This paper presents the results of a joint effort between the ALFI Risk Management working group and ALRiM on the backtesting of Value-at-Risk (VaR) models, i.e. the "means of examining whether or not reported VaR represents an accurate measure of [...] the actual level of risk" (Campbell, 2005). It proposes practical guidelines on how to perform and interpret backtests. This will help practitioners extract more value from VaR models and better understand the market risk of their UCITS. Given this scope, this paper is intended for risk managers and conducting persons in charge of risk management and adopts a somewhat technical stance.

The first section of this paper provides a definition of backtesting, as well as a summary of key aspects to consider. This section also contrasts model validation with backtesting, compares the use of hypothetical and actual P&L statements (clean and dirty backtesting), explains the types of error that can arise in backtesting, and defines the power of a test.

The second section describes a set of elementary backtesting techniques. The two basic properties a VaR model should jointly satisfy are unconditional coverage and independence. Unconditional coverage refers to the fact that outliers (occurrences of losses worse than those predicted by the VaR) should not significantly differ from the theoretical proportion (typically 1% under Luxembourg regulation). The independence property means that outliers should occur independently of each other. Several statistic tests aiming to test these properties are presented.

The third section lists analysis procedures that need to be included in ongoing backtesting reviews and areas that could be investigated on an ad-hoc basis. Interpreting the outcome of these tests is of utmost importance. These tests cannot and should not replace or supersede a full-blown analysis and understanding of models: instead, they are to be considered as tools allowing the user to reduce the dimensionality of a complex problem and therefore eventually raise red flags when warranted.

Appended to the paper, the reader will find a references section for further reading, an exhaustive glossary of terms used in this paper, and specific formulas and explanations to implement the tests described in Section II including hypothesis statements, test statistics, and decision rules.





# Introduction

This paper presents the results of a joint effort between the ALFI Risk Management working group and ALRiM on the backtesting of Value at Risk (VaR) models. The first of a three-paper series, it describes the conceptual and practical basics of backtesting – how to properly backtest a VaR model beyond the obvious steps mandated by UCITS regulations. Future papers, in turn, will discuss how to respond to the apparent failure of a VaR model and how to structure related governance arrangements.

VaR is commonly used to calculate the exposure to market risk, called global risk exposure in the UCITS context. The financial crisis and subsequent events, however, have highlighted issues concerning the calibration and reliability of VaR models. The industry has questioned not only the use of the VaR, but also its accuracy and its ability to capture the risk. Meanwhile, the CESR guidelines 10-788 (especially Box 18), the CSSF Circular 11/512 as well as article 45(2)b of CSSF regulation 10-4 require monthly backtesting for UCITS calculating their global risk exposures using a VaR model and provide additional guidance.

This paper, then, proposes practical guidelines on how to properly perform and interpret backtesting. It is structured in four sections. Starting with definitions and basic distinctions, it then describes some elementary statistical backtesting techniques before proposing additional, complementary analysis tools. The concluding section points to additional considerations to improve the backtesting framework.

> "Disclosure of quantitative measures of market risk, such as value-at-risk, is enlightening only when accompanied by a thorough discussion of how the risk measures were calculated and how they related to actual performance"
>
> Alan Greenspan (1996).

## Why backtest VAR models?

Value-at-Risk (VaR) measures are part of the broad array of financial indicators that practitioners use and rely on daily for risk monitoring and decision making. Unlike most other indicators, a VaR measure is not a descriptive statistic that can be unequivocally "measured": statistically speaking, VaR is a forecasting measure[2] provided by a selected model relying on a set of underlying assumptions.

Therefore, it is merely estimated by one of the many financial models dedicated to this task. These VaR models sometimes yield diverging predictions for the same portfolio at the same date. This fact is well-known not only among risk managers, but also among the portfolio managers and traders whose freedom is constrained by VaR limits. Hence the need to demonstrate the accuracy of the forecasts provided by a given VaR model with respect to the actual subsequent observations.

This demonstration can never be done once and for all. VaR models rely, explicitly or not, on a set of assumptions that reasonably reflect real market dynamics most of the time. But sometimes these dynamics change, the assumptions become less reasonable, and the models appear to fail: for instance, portfolio losses start exceeding the VaR with alarming regularity.

Thus, the performance of the VaR model against portfolio returns must constantly be reassessed to maintain the VaR's credibility. This assessment is the primary purpose of backtesting, defined by Campbell (2005) as the "means of examining whether or not reported VaR represents an accurate measure of [...] the actual level of risk".

Beyond making the VaR a more authoritative indicator, backtesting also helps strengthen the market risk measurement process. Recall that some VaR models are based on distributional assumption (e.g. normality, independent and identically distributed (i.i.d.) returns, etc.) while others assume that past returns, as defined in a certain way, closely approximate future returns. Model users always need to be aware of such assumptions and of the limitations of their models. They need to regularly question and test whether these assumptions are verified by the facts. Checking how the model performs in practice leads to examining the suitability of the risk measurement models and their assumptions to find possible shortcomings. This, in turn, helps improve the risk measurement process.

[2] VaR is an application of the broader concept of distribution quantile forecasting applied in the context of risk management.



## Backtesting vs. Model Validation

Model validation being a topic in its own right, this section is voluntarily kept short and aims to contrast backtesting and model validation.

An independent validation is mandatory as per Box 22 of the CESR guidelines 10-788 following the initial development (or any subsequent significant change[3]) of a VaR model to calculate the global exposure of a UCITS. The goal is to ensure that the model is conceptually sound and captures adequately all material risks. This independent validation has a broader perspective than backtesting: it also covers the scope, specification, implementation and use of the model.

Validation would for instance need to answer the question: is the model appropriate for the type of portfolio, the type of data, and the systems in place? The (internal or external) party performing the validation must be independent of the model building process.

Backtesting, on the other hand, is a quality control process that does not primarily deal with how VaR numbers were generated, but focuses more narrowly on one question: is the model able to accurately reflect the observed reality? It is only when the answer is negative that backtesting raises questions relating to model validation.

CESR guidelines further specify: in addition to this initial independent validation, Risk Management should perform 'ongoing validation of the VaR model (this includes, but is not limited to back testing […]) to ensure the accuracy of the model's calibration. The review should be documented. Where necessary, the model should be adjusted'.

In short, backtesting is part of (and the first step in) the ongoing validation of the model by Risk Management, but not necessarily of the independent initial model validation.

## Clean/dirty backtesting

Backtesting is called:

- Dirty when it compares the VaR measures with effectively observed returns, and
- Clean when it compares the VaR with hypothetical historical returns, i.e. the daily returns that would have been observed for each day if portfolio positions had been kept unchanged from the previous day.

Both backtesting approaches are accepted by the CSSF and ESMA. The explanatory text included in ESMA guidelines (item 58) states that 'back testing is ideally performed on the hypothetical changes in the portfolio's value' (i.e. clean backtesting). In practice, most UCITS have a relatively low portfolio turnover compared to the bank trading desks for which the VaR was originally developed. Given this, the additional accuracy and insight that can be gained from clean backtesting is rarely perceived to justify the additional costs of computing and maintaining hypothetical returns, so that dirty backtesting is widely accepted[4].

[3] The CESR guidelines 10-788 specify in Box 22, point 3 that a significant change could relate to the use of a new product by the UCITS, the need to improve the model following the back testing results, or a decision taken by the UCITS to change certain aspects of the model in a significant way.

[4] An arguable theoretical superiority of clean backtesting needs to be counterbalanced by the technical difficulties implied. Nevertheless, in some specific occurrences (e.g. for portfolios with a high daily turnover), clean backtesting might be the only route to explaining a model overshooting.

## Counting the hits

Whether they are performed on clean or dirty data, most basic backtesting procedures start with a counting exercise. This first step is the only one specifically required by UCITS regulations. Recall that VaR estimates the (worst) potential (mark-to-market) loss at a given confidence level (probability) over a specific time horizon[5]. Whenever the daily P&L is worse than the VaR figure computed for that day, we observe a "hit" (also called exception, outlier, exceedance, violation or overshooting). The usual backtesting process starts with counting these.

A hit is not worrying by itself: a correctly calibrated VaR 99% model generates about 1 hit every hundred observations and, on average, 2.5 per year (250 observations for daily observations). But given that our testing sample is subject to random fluctuations, we will often observe less than 2 or more than 3 hits over the past year, even if our VaR model is correct. In fact, given a correctly calibrated VaR 99% model and 250 observations, we can expect to count hits with the following probabilities[6]:

| Number of hits | 0 | 1 | 2 | 3 | 4 | >4 |
|---|---|---|---|---|---|---|
| Likelihood | 8.1% | 20.5% | 25.7% | 21.5% | 13.4% | 10.8% |

For UCITS calculating their VaR at the 99% confidence level, the CESR guidelines and the CSSF circular indicate 4 hits in the most recent 250 business days as the critical threshold. Above 4 hits, a report must be submitted at least quarterly to senior management, containing an analysis and explanation of the exceptions and a statement of remedial actions, if any. As the table above makes clear, even if the VaR model is correct for the portfolio where it is applied, this threshold will be exceeded more than 10% of the time[7]. This helps explain why discussions on this topic recur with such regularity.

The painfully frequent handwringing about VaR exceedances explains the growing consensus in the industry: backtesting cannot be limited to the basic counting exercise required by regulations. Fortunately, in recent decades the academic literature has developed a wide array of tests based on statistical inference.
While these statistical tests are too many and too complex to present them all here, the most usual and widespread are described in the following section.

[5] By some definitions, this estimation should correspond to 'normal market conditions'.
[6] The theory behind this table is explained in Section II under 'Unconditional Coverage Tests'.
[7] Some UCITS use a VaR 95% model, implying one hit every 20 observations on average. Assuming that the VaR model is correct, they will observe more than 16 hits about 12.5% of the time and more than 17 hits about 8% of the time. These UCITS should arguably be allowed to apply a >16-hit threshold for compliance with the CESR guidelines.



# II. Statistical backtesting techniques

Statistical tests are used to assess the quality of VaR measurements in a finite sample inherently subject to randomness. Most of them qualify the number and/or pattern of exceptions. This allows their users, beyond regulatory requirements, to define acceptable zones and to raise red flags when the model performs outside of these zones. Some caveats are in order at this point:

First, test results need to be completed by further analysis: for instance, the reasons behind the occurrence of outliers need to be investigated and documented. We cover this in the next section.

Furthermore, the tests listed in this section are just examples that may be used as part of a more comprehensive toolbox. They are all 'model-free' in the sense that they apply to any type of VaR model, without directly addressing the models' assumptions. Despite their limitations, they provide a good starting point to a more comprehensive analysis of backtesting.

Finally, only the main drivers of the tests are reviewed here, in order not to overload the paper with statistical considerations, notations and details. For interested readers, mathematical formulas are presented in the annex, and comprehensive descriptions and reviews are included in the referenced papers.

## When to reject a model?

All statistical backtests ask the same question: given a certain significance level, should we reject the "null hypothesis" that our model adequately forecasts the losses distribution, or exhibits a specific property?

The choice of significance level is related to an assessment of the costs of making two types of mistakes: we could reject a correct model (type I error) or we could fail to reject an incorrect model (type II error).

| | | Model: Correct | Model: Incorrect |
|---|---|---|---|
| **Decision** | **Not reject** | OK | Type II error |
| | **Reject** | Type I error | OK |

There is a trade-off between the two: increasing the significance level generates more type I errors but less type II errors, and vice versa. In risk management, type II errors may be very costly, so that a significance level of 10% may be appropriate.

This seems to match CESR's view: as mentioned above, for a correctly calibrated VaR 99% model, the probability of observing more than 4 hits (i.e. 5 hits or more) in the last 250 days is 10.8%.

## How powerful is your test?

Ideally, we want to minimise the likelihood of both types of errors to obtain a more powerful test. The simplest way to get there is to work with more data by:

- Using the longest possible P&L series (ideally since the last material change to the model), and/or
- Decreasing the VaR confidence level: this will generate more hits and therefore more powerful tests (but without alleviating the need to count the hits vs. VaR 99% for regulatory purposes).

The other way to perform more powerful tests is to use procedures that have been shown to more accurately separate correct from incorrect VaR models. This comes at a cost: more powerful tests are usually more complex.

## Backtesting properties

In a much-cited paper, Christoffersen (1998) explained for the first time what is now considered obvious: any model to forecast intervals, such as a VaR model, is correct if the exceptions it generates are not only quantitatively in line with its confidence level, but also independent of each other. In other words, a VaR model should jointly satisfy two conditions:

- Unconditional coverage (UC): the observed proportion of hits (coverage) should not be significantly different from the theoretical proportion, e.g. typically 1% under Luxembourg regulation. Significantly more frequent exceptions suggest that the reported VaR figures systematically understate the fund's actual level of risk. Conversely, too few violations suggest an overly conservative VaR measure.

- Independence (IND): This property concerns not the frequency, but the sequence of outliers. The fact that an outlier is observed today (or not) should not affect the probability of observing an outlier on the following days.

In other words, any two elements of the hit sequence must be independent of each other. Breaches of this property should influence our assessment of risk. For instance, if the backtest of the 99% VaR yields 1% outliers, but these are concentrated on a two-week period, then the model is less responsive to market changes (and the portfolio may be much riskier) than a correct model generating violations that occur randomly through time. Therefore, VaR models which generate violation clusters should be rejected.

Campbell (2005) points out that the unconditional coverage and independence properties jointly determine the accuracy of a given VaR model. A VaR model that satisfies only one property or the other will result in an inaccurate description of the risk exposure. Accordingly, most statistical procedures used in backtesting serve to test the independence or the unconditional coverage property, or even both, as described below.

## Elementary statistical tests

Most statistical methods used in practice examine the properties presented above. They provide a metric to evaluate assertions such as 'The frequency of outliers is close to 1%' and 'Outliers are distributed independently'.

To illustrate the need for a metric, consider Fund A and Fund B monitored with VaR 99%. Here are the results of the last 320 observations:

**Fund A**

| | |
|---|---|
| Number of observations | 320 |
| Number of outliers T1 | 5 |

**Fund B**

| | |
|---|---|
| Number of observations | 320 |
| Number of outliers T1 | 18 |

For both funds, the rates of outliers (at 1.6% and 5.6% respectively) are above the 1% threshold corresponding to the 99% confidence level. Just how high is too high? This is the main question addressed by unconditional coverage tests.

They examine whether the number of exceptions observed in the testing sample are in line with the confidence level chosen for the VaR.

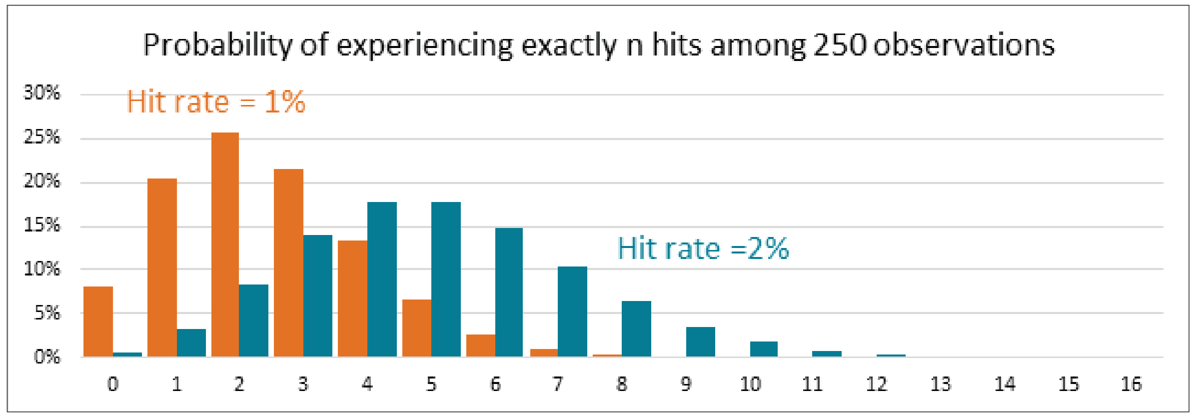


## Unconditional coverage test – one-tailed test

To test the unconditional coverage of a VaR 99% model, the most intuitive approach is to consider each trading day as a Bernoulli trial with a 1% probability of a hit. Then, if the model is correct, the number of hits follows a binomial distribution, characterised by a hit rate of 1% and the number of observations in the sample.

Now let's assume that, as is often the case, we are only concerned about "too many", but not "too few" exceptions to the VaR. In other words, we only want to conduct a "one-tailed" test. Then our null hypothesis is that the number of hits is not significantly higher than 1% times the number of trading days in the sample. The complete test procedure is described in Annex 1. In a nutshell, we first define the highest acceptable number of hits given the sample size (and VaR confidence level, i.e. 99% in our case), and then compare it to the observed number of hits.

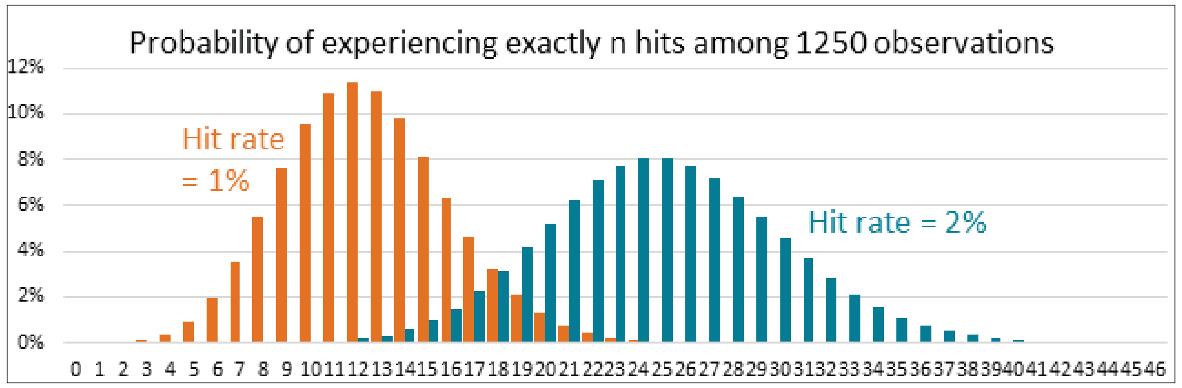
Following this procedure, we quickly find that using only 250 observations and a 99% confidence level, "more than 5 observations" should lead to rejecting the null hypothesis of adequate unconditional coverage, because for a VaR model with a 1% hit rate this should happen about 10.8% of the time. But if we decide not to reject it, how confident can we be that this decision is the right one? Not very, as the following graph exemplifies:




Here we clearly see that if our one-year data contains less than five hits (i.e. we will not reject the null hypothesis), especially if it contains three or four hits, we are still far from sure that our VaR model has a 1% hit rate. It may in fact have, for example, a 2% hit rate instead. In other words, it may in fact be a VaR 98% model! If that were the case, observing less than five hits would occur with a 44% probability[8].

Hence, a non-rejection should not really reassure us, because the default VaR confidence level and number of observations mentioned by regulations lead to a test of unconditional coverage that lacks in power.

Compare this with the situation when using 5 years of data, i.e. about 1250 observations:




[8] Or a VaR model with a 3% hit rate (13% probability) or with a 4% hit rate (3% probability), etc.

Here, using the procedure in Annex 1, we calculate 16 as our highest acceptable number of hits. Now the probability of observing so few hits for a VaR model with a 2% hit rate is only about 4%. That is a lot less than 44%. This simple example illustrates how using more data may considerably reduce the probability of a type II error. This is also clearly visible from the much lower degree of overlap between the two distributions in our second chart.

Hence the advantage of testing the VaR model using the longest possible data history in addition to regulatory requirement (most recent 250 days). This, however, should not lead to the inclusion of data from a previous version of the model! The loss in backtesting power is part of the cost of significantly amending VaR models.

Extending the historical window does though come with costs: when running or testing risk numbers, the risk manager needs to find an adequate balance between including additional data in order to improve the representativeness of the sample, keeping data representative of current market conditions (changes to the model, capturing emerging trends or patterns, etc.) and the burden of maintaining and manipulating larger data sets.

As increasing the data set could reduce the risk of some errors by e.g. making the sample reviewed more representative of the population, it might, if unduly handled, increase the probability of other errors by masking the impact of changes (model, recent trends, etc.).

For models with short histories, an alternative is to also compute the VaR 95% and to conduct backtests on their time series in addition to the VaR 99% series.

Let's take 10% as our significance level and get back to Fund A and Fund B. Using the procedure in Annex 1, we find that for 320 observations, the highest acceptable number of hits is 5, exceeded only 10.4% of the time.

For fund A, the number of hits is just at the upper end of the acceptable range: we do not have sufficient evidence to reject the null hypothesis that the VaR model is in line with the 99% confidence level. Failing to disqualify the model accuracy leads us in practice to accept it until evidence to the contrary emerges.

No so for fund B, however. 18 hits is clearly out of range. In fact, pursuing the same procedure until 18, we find only a 6.8 x 10-9 probability of observing 18 or more hits beyond the VaR 99% among 320 observations if the model were correct. Therefore, we can reject the null hypothesis that our VaR model provides adequate unconditional coverage.

### Unconditional coverage test – two-tailed tests

Given enough data on daily returns, it is also possible to test whether a VaR model generates not just too many hits, but more generally a quantity of hits that is not in line with its confidence level: too many OR too few hits. Such a 'two-tailed' test allows its user to estimate whether the VaR model systematically underestimates or overestimates risk. In this case, the 10% significance level must be split between the two tails (5% on each side) and the procedure in Annex 1 must be amended accordingly.

Kupiec (1995) provides a more elegant, widely used two-tailed test. Here we examine whether the hit rate, called "proportion of failures" or POF, is in line with the VaR model's α (one less the VaR confidence level).

The test statistic is based on a log-likelihood ratio, an approach which has been shown to be especially powerful to distinguish between two models with no unknown parameters. The null hypothesis, test statistic and decision rule are presented in Annex 2.

### Test of unconditional coverage over a range of alpha levels – Pearson´s Q test

Tests of unconditional coverage performed for a single confidence level (i.e. a single point of the return distribution) are not very powerful for the small samples typically used in VaR backtesting. As we have seen above, the risk of type 1 or type 2 error is quite high – especially when the VaR confidence level is high and the number of hits accordingly low. Therefore, tests that simultaneously cover multiple confidence levels have been proposed. CSSF also guided towards testing various confidence intervals in its 2014 annual report. Pearson's Q test requires users to divide the unit interval into different sub-intervals, such as [0.00,0.01[, [0.01,0.05[, [0.05,0.10[, [0.10,1.00] for instance, considering that we are only interested in measures over the 10th percentile.

This results in four separate bins on the unit interval. After choosing the partition, we count the number of VaR violations that occur within each bin. For example, the [0.01,0.05[ bin records the number of days on which a loss occurred that exceeded the 5% VaR, but not the 1% VaR. Given the number of hits that have occurred within each bin, the test statistic is calculated as detailed in Annex 3, which also provides a decision rule.

In his review of VaR models, Campbell (2005) explains that tests that examine multiple quantiles exhibit a higher detection power most clearly when applied to VaR models that understate the risk most severely.



### Independence testing

A model can continue to perform handsomely in tests of unconditional coverage while the fund is experiencing a string of VaR violations that spiral into catastrophic losses. We therefore need a test that will reject a VaR model generating clustered violations[9].

Christoffersen (1998) focuses on unusually frequent consecutive hits. His Markov test[10] examines whether the likelihood of a VaR violation depends on a VaR violation being observed on the previous day. If the VaR measure is correct, then the probability of violating today's VaR should be independent of whether yesterday's VaR was violated or not. As noted by Campbell (2005), if the likelihood of a VaR violation increases on days following previous VaR violations, then the VaR directly after a violation should be increased.

The Markov test is carried out by creating a 2x2 contingency table that records VaR violations on consecutive days:

| | $I_{t-1}=0$ | $I_{t-1}=1$ | |
|---|---|---|---|
| $I_t=0$ | $N_1$ | $N_2$ | $N_1+N_2$ |
| $I_t=1$ | $N_3$ | $N_4$ | $N_3+N_4$ |
| | $N_1+N_3$ | $N_2+N_4$ | $N$ |

Where It is a binary function of value 1 if an outlier is recorded at date t, and 0 otherwise. This function is often referred to as the 'hit function'.

The proportion of violations on the days that follow a day when no violation occurred, $\pi_0 = N_3/(N_1+N_3)$, should be the same as the proportion of violations that immediately follow another violation, $\pi_1 = N_4/(N_2+N_4)$. Thus, our null hypothesis is $\pi_0 = \pi_1$.

To perform this test, we also define our proportion of violations (hit ratio) as $\pi = (N_3+ N_4)/(N_1+N_2+N_3+N_4)$, and then compute our test statistic, just like in Kupiec's POF test, as a likelihood ratio. The statistic and associated decision rule are presented in Annex 4.

Let's illustrate again with Funds A and B.

**Fund A**

| | |
|---|---|
| N1 | 310 |
| N2 | 5 |
| N3 | 5 |
| N4 | 0 |

For fund A, no sequence of outliers is observed. Since $N_4 = 0$, $\pi_1 = 0$ as well, so that the denominator of the likelihood ratio reduces to $(1 - \pi_0)^{N1} \pi_0^{N3}$. As expected, the test statistic (at 0.1587) is below the critical value for any of the usual significance levels. Therefore, the hypothesis of independence cannot be rejected.

**Fund B**

| | |
|---|---|
| N1 | 292 |
| N2 | 10 |
| N3 | 10 |
| N4 | 8 |

[9] Which might be particularly relevant for financial variables that exhibit cyclicality and time changing risk levels.
[10] To be exact, we should say 'Test of the transition probabilities of the Markov chain'. We opted for the simpler but not totally accurate 'Markov test' formulation.

Unlike Fund A, Fund B displays 8 instances of consecutive outliers. At 26.02, the test statistic is an extreme value of its probability distribution under the null hypothesis, exceeded only with a probability of 3.4 x $10^{-5}$ %. Therefore, we can reject the null hypothesis of independence at any of the usual significance levels.

### Further Tests of Independence

An obvious shortcoming of the Markov test is that it will not detect hit clusters that contain no, or only a few, consecutive hits. To address this, Christoffersen & Pelletier (2004) propose another test of independence, based on counting the days between hits (a measure called "duration"). The null hypothesis here is that, given a correctly specified VaR model with coverage rate α, the expected duration until the next hit should be a constant 1/α days, whether the last hit occurred recently or in the distant past. Compared to the Markov test, however, the original duration-based test of independence by Christoffersen & Pelletier involves more sophisticated, less transparent computations which make its use in risk reporting problematic. This is also true for the other duration-based tests of independence that have been proposed so far by academics.

Other approaches to testing for independence include testing whether the hit function verifies the properties of a martingale difference (Hurlin & Tokpavi, 2006), or whether current violations can be linked by linear regression to past violations (Engle and Manganelli, 2004). Like duration-based tests, they offer higher power at the cost of higher complexity and opacity.

### Joint test of Unconditional Coverage and Independence

Further to the Markov test for independence, if the VaR measure also exhibits the unconditional coverage property then the proportions of violations should be equal not only to each other, but also to the "hit rate" and to the likelihood associated with the $VaR_{1-\alpha}$ confidence level (e.g. α=1% for VaR 99%). In other words, our null hypothesis here is:

$$N_3/(N_1+N_3) = N_4/(N_2+N_4) = (N_3+N_4)/N = \alpha.$$

The test statistic for this test is simply the sum of the test statistics for the Kupiec's POF test and Christoffersen's Markov test, as defined in annex 2 and annex 4. The decision rule for this test is explained in annex 5.

There is a hidden cost to testing both properties simultaneously: for reasons explained in annex 5, a good unconditional coverage could mute the effects of a lack of independence or vice-versa.
As Campbell (2005) puts it, "The fact that one of the two properties is satisfied makes it more difficult for the joint test to detect the inadequacy of the VaR measure". Therefore, standalone tests of unconditional coverage and independence must also be performed whenever a VaR model passes the joint test.

Following our examples, we can show that the test statistics is high enough in the case of Fund B to reject the null hypothesis that the model has adequate unconditional coverage and independent hits. For fund A, on the other hand, the test statistics are too low to reject the null, so that further separate tests of independence and unconditional coverage are needed.

### Further types of tests[11]

One characteristic of all the tests presented above is that they give every exceedance the same weight, whether it was very close to the VaR or wiped out a significant portion of the fund. The size of exceedances is just irrelevant in the definition of the VaR. In the words of hedge fund manager David Einhorn, a 99% Value at Risk calculation does not evaluate what happens in the last one percent… This is like an air bag that works well all the time except when you have an accident.

To help alleviate the unease on this topic, Lopez (1998) developed a test that considers the size of exceedances. The practical problem when implementing it is to decide how large a loss is too large.

Finally, distribution tests apply to the VaR methodologies that are based on a distributional assumption, and test whether this assumption is verified at all levels of the distribution, not just the 95% or 99% level. Unlike all the "model-free" methods presented above, these tests are model-specific.

The latest developments in VaR models' performance evaluation have been focused on comparing alternative VaR models rather than evaluating the models' absolute performances individually. These new generation tests allow to identify the "best" VaR model with a given confidence level. For that purpose the tests mainly rely on loss functions specially designed to account for the frequency and magnitude of exceptions.

[11] An exhaustive review of the existing VaR testing procedures can be found in Nieto and Ruiz (2016).



While these tests are not a substitute for more standard backtesting procedures, they are useful tools to select or benchmark competing VaR models.

### Conclusion

Testing for both unconditional coverage and independence will result in one of the following outcomes:

- Reject (the null hypotheses for) both properties,
- Reject one property, but not the other, or
- Not reject any of these properties.

The two former outcomes warrant further probing and analysis: either the model is flawed and needs to be adapted/enhanced as it doesn't assess risk properly, or some other implicit assumption has been breached.

The fact that a model failed the backtesting procedures doesn't necessarily mean that the model is faulty and should be replaced. Most probably, such an outcome points to a poorly specified or even a mishandled model needing some tailoring to specific situations.

In the next section, we offer some avenues of reflection and analysis that could be exploited to foster a better understanding of the practicalities of the VaR model used. A thorough understanding of this model allows the risk manager to better grasp its limitations and the improvements required when the backtesting process flags potential model weaknesses or misuses.

## III. Beyond statistical analysis

As suggested in Section I when defining backtesting, users should understand the limitations inherent in each model.

The statistical approaches outlined above will help to raise red flags. Pinpointing the exact problem, however, will require the user to delve further into the details and practicalities underlying the model. This approach needs to be further formalised via the construction of a comprehensive toolkit allowing reality checks on all aspects and (explicit and implicit) modelling assumptions.

Such a toolkit will uncover problems and allow its users to plan for corrective measures, or at least factor any identified bias into the decision-making process. The goal is to set up indicators allowing model users to adapt these models before they fail.

The first areas to investigate on an ongoing basis include:

- Review moves of the underlying risk factors;
- Check for large changes in positions;
- Probe for event risk that might not have been captured (market-impacting events such as FED tapering, Lehman bankruptcy, Fukushima meltdown, etc.); and
- Contrast backtesting results to expected shortfall and stress-test results to gauge result plausibility.

Additional areas of ad-hoc investigation may include:

- Depending on the VaR model, it may be possible to use tailored backtesting procedures that examine the model's underlying assumptions (unlike the "model-free" backtesting methods described here).
- Graphical checks, illustrating for instance the timing of VaR exceedances to check for any form of clustering not identified by the Christoffersen test described above.
- In addition to VaR results, risk managers need to pay attention to what happens in the tail. For instance, very large exceedances need to be given attention, especially since the highest VaR limit utilisation needs to be disclosed as part of the annual risk report.
- Backtesting results at the industry level could be considered to establish (dynamic) policies and limits.

Finally, the exclusive focus on "too many" exceptions may hide a whole set of VaR models generating too few exceptions. The absence of exceptions also need to be considered by risk managers to revisit their models.



## IV. Final comments

Throughout this paper, we aimed to provide elementary information on how to define and set up a practical backtesting framework. We voluntarily restrained our scope to the technical and practical issues and did not cover:

- How to respond to the apparent failure of a VaR model, especially when too many exceptions occur;
- The governance arrangements that backtesting need to be embedded in.

Both aspects will be covered in upcoming papers that will round out this series on VaR backtesting. Please note that ALFI-ALRiM, while confident of the broad aims and principles highlighted in this white paper, does not endorse the practical use and specific implementation of the described principles and tools or associated spreadsheets. It is strongly advised to fully review and test the material before using it for professional purposes. ALFI, ALRiM and group members therefore cannot be held liable for any error, imprecision or mistake endured following the use of the provided material.

## Bibliography

Documents referenced in this paper


Basel Committee for Banking Supervision (1996), Supervisory Framework for the Use of « Backtesting » in Conjunction with the Internal Models Approach to Market Risk Capital Requirements

Campbell S. (2005), A Review of Backtesting and Backtesting Procedures

Committee of European Securities Regulators (2010), CESR's Guidelines on Risk Measurement and the Calculation of Global Exposure and Counterparty Risk for UCITS, CESR/10-788

Christoffersen P. (1998), Evaluating Interval Forecasts

Christoffersen P., Pelletier D. (2004), Backtesting Value-at-Risk: A Duration-Based Approach

Commission de Surveillance du Secteur Financier (2011), CSSF Circular 11/512

Engle R., Manganelli S. (2004), CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles

Hurlin Ch., Tokpavi S. (2006), Backtesting VaR Accuracy: A New Simple Test

Kupiec P. (1995), Techniques for Verifying the Accuracy of Risk Management Models

Lopez J. (1998), Methods for Evaluating Value-at-Risk Estimates

Nieto M.R., Ruiz E. (2016), Frontiers in VaR forecasting and backtesting


Further reading


Berkowitz J. (2000), Testing Density Forecasts with Applications to Risk Management

Berkowitz J. (1998), Evaluating the Forecasts of Risk Models

Berkowitz J., O'Brien J. (2001), How Accurate Are Value-at-Risk Models at Commercial Banks

Berry R., JPM Investment Analytics and Consulting (Newsletter Sept 2009), Backtesting Value-at-Risk

Boudoukh J., Richardson M., Whitelaw R. (1998), The best of Both Worlds: A Hybrid Approach to calculating Value at Risk

Candelon B, Colletaz G et al. (2009), Backtesting Value-at-Risk : A GMM Duration-based Test

Christoffersen P. (2002), Elements of Financial Risk Management, Chapter 8 – Backtesting and Stress testing

Christoffersen P. (2006), Value-at-Risk Models

Commission de Surveillance du Secteur Financier (2007), CSSF Circular 07/308

Hurlin Ch., Tokpavi S. (2007), Une Evaluation des Procédures de Backtesting

Jorion Ph. (2009), Risk Management Lessons from the Credit Crisis

Jorion Ph. (2007), Value-at-Risk: The New Benchmark for Managing Financial Risk

Pena V, Rivera R, Ruiz-Mata J. (2006), Quality Control of Risk Measures: Backtesting VaR Models

Pritsker M. (2001), The Hidden Dangers of Historical Simulation

Zangari P (1996). An improved methodology for measuring VaR




**Backtest** — Process of backtesting

**Backtesting** — Formal statistical framework that consists in verifying if actual trading losses are in line with projected losses. This involves a systematic comparison of the history of model-generated VaR forecasts with actual returns.

**Christoffersen test** — This statistical test examines a version of the independence property, whereby independence is only assessed by reference to consecutive VaR violations.

**Clean backtesting** — Use of the hypothetical changes in the portfolio's value for backtesting. Hypothetical changes are the changes in value of the portfolio from one day to the next assuming unchanged positions.

**Clusters** — Instance of multiple VaR violations occurring within a tight time interval.

**Conditional coverage (CC)** — CC = UC + IND

Conditional coverage requires that both properties of unconditional coverage and independence be jointly validated.

Further to the average number of exceptions being correct, these occur independently throughout the observed sample. This means that for any given day, the probability of occurrence of an exception is the same, whether an exception occurred previously or not.

**Cornish-Fischer transformation** — Technique for approximating an empirical distribution by integrating not only the information conveyed by the mean and standard deviation of the distribution but also its skewness and kurtosis.

**Dirty backtesting** — Use of the effective changes in the portfolio's value for backtesting. Effective changes are the changes in value of the portfolio from one day to the next (including the impact of transactions).

**Error**

- **Type I error** — Rejecting a correct model
- **Type II error** — Not rejecting an incorrect model

The choice of significance level is related to an assessment of the costs of making two types of mistakes: rejecting a correct model (type I error) or accepting an incorrect model (type II error). There is a trade-off between both types of error, since increasing the significance level implies more type I errors but less type II errors, and vice versa. In risk management, type II errors may be very costly so that a significance level of 10% may be appropriate.

**Exception** — One-day change in the portfolio's value that exceeds the related one-day value-at-risk measure.



**Exponential weighting** — Technique used to give more weight to recent information than to older information. Typically, historical VaR uses equally-weighted P&L's whereas parametric methodologies tend to use weighted P&L's. The idea between this weighting is that recent information conveys more relevant information for today's risk measure.

**Global exposure** — Global exposure is a measure designed to limit either the incremental exposure and leverage generated by a UCITS through financial derivatives instruments (including embedded derivatives), or the market risk of the UCITS portfolio.

Authorized measures of global exposure include the commitment approach and the (absolute or relative) value-at-risk approach.

**Heteroscedasticity** — Refers to the fact that the variance of a random variable is not constant or that the variances of several random variables differ. The presence of heteroscedasticity might induce bias in models when trying to project future values (statistical inference). This is one of the underpinnings of using daily returns (or returns of non-overlapping periods) in VaR computation.

**Hit** — See 'Exception'

**i.i.d.** — Short for independent and identically distributed

**Independence (IND)** — This refers to the capacity of the model to integrate information so that outliers do not cluster together.

In Christoffersen's 1998 version of the concept, the independence of two consecutive elements of the hit sequence means that the probability of observing an outlier on any day should not be affected by the occurrence (or not) of an outlier on the previous day.

**Independent and identically distributed returns** — Each series of return has the same probability distribution as the others and all are mutually independent. This widely-used assumption simplifies the aggregation of series (foundation of the central-limit theorem) and allows scaling (square root of time rule). Such a process is typical of e.g. multiple draws using a roulette or a stochastic Wiener process.

This assumption is not empirically verified. Although it might be reasonable for consecutive daily returns, it gets shakier as the analysis horizon is expanded (among others, because returns exhibit heteroscedasticity).

**Inference statistical testing** — Testing of a hypothesis through procedures used to draw conclusions from datasets arising of random variables.

**Kupiec test** — This statistical test focuses exclusively on the property of unconditional coverage (UC), testing if the reported frequency of VaR outliers is significantly more or less than the theoretical proportion.

| | |
|---|---|
| Kurtosis | Measure of the 'peakedness' of the probability distribution. Higher kurtosis means that more of the variance is the result of infrequent extreme deviations. This is the moment of order 4 of a distribution. A normal distribution has a kurtosis of 3 or excess kurtosis of 0. |
| Leptokurtic | A distribution with (positive) excess kurtosis. In terms of shape, a leptokurtic distribution has a more acute peak around the mean (that is, a lower probability of values near the mean) than a normal distribution and fatter tails (that is, a higher probability of extreme values than a normal distribution).<br><br>Leptokurtic series are also known as being 'fat-tailed'. |
| Mapping | Mapping is understood as the process of modelling the behaviour of a given instrument by a restricted amount of risk factors. For instance, this could be the definition of a yield curve using a finite number of points and the specification of an interpolation methodology for intermediate maturities. |
| Markov test | The Markov test is used to test the independence (IND) property of VaR models. It examines whether the likelihood of a VaR violation depends on a VaR violation being observed on the previous day. The idea is to establish a test which will be able to reject VaR with clustered violations. |
| Model validation | Validation has a broader perspective than backtesting in the sense that validation will cover additionally the scope and use of the model. Validation would for instance need to answer to the question: is the model appropriate for the type of portfolio, the type of data and the systems in place? |
| Null hypothesis | The null hypothesis typically corresponds to a general or default position. For example, the null hypothesis might be that there is no relationship between two measured phenomena. It is important to understand that the null hypothesis can never be proven. A statistical test can only reject a null hypothesis or fail to reject it.<br><br>E.g. for backtesting, the null hypothesis is usually that the model is correct. Rejecting the null hypothesis leads the risk analyst to suppose that the model is not correct. |
| Outlier | See 'Exception' |
| Overshooting | See 'Exception' |
| POF | Proportion of failures. Kind of statistical test used to test an assumption such as adequate coverage. Kupiec is part of this family of tests. |
| Reactivity | Capacity of a model to incorporate the most recent market information in its predictions. |
| Sampling bias | The selection of a given sample on which to base statistical measures or inference analysis could induce a bias as this sample may not be representative of the population and therefore "real" expected values.<br><br>This is highly relevant for VaR computations as all computations are based on some sort of sampling. Furthermore, users need to allow a trade-off between increasing the sample size to achieve statistical relevance and making optimal use of the most recent and relevant information by reducing the sample size. |
| Skewness | Measure of the asymmetry of the probability distribution. This is the moment of order 3 of a distribution.<br><br>Qualitatively, a negative skew indicates that the tail on the left side of the probability density function is longer than the right side and the bulk of the values (possibly including the median) lie to the right of the mean. A positive skew indicates that the tail on the right side is longer than the left side and the bulk of the values lie to the left of the mean. A zero value indicates that the values are relatively evenly distributed on both sides of the mean, typically but not necessarily implying a symmetric distribution. A normal distribution has a skew of 0. |
| Unconditional coverage (UC) | This refers to the capacity of the model to correctly predict the probability of a VaR violation: the fraction of outliers should not be significantly different from the theoretical proportion (e.g. typically 1% under Luxembourg regulation). |
| Value-at-Risk | A measure of the potential loss to a portfolio due to market risk. More specifically, VaR measures the potential mark-to-market loss at a given confidence level (probability) over a specific time horizon (assuming, under certain definitions, "normal market conditions". |
| VaR | Short for Value-at-Risk |
| Violation | See 'Exception' |

## Annex 1

### Procedure for a one-tail test using the binomial distribution

1) Choose a significance level for the test (unrelated to the VaR confidence level). The more data is available, the more powerful the test, and the less the significance level needs to be.

2) For each integer x, starting from 0:
   a) Compute the probability of experiencing x exceptions if the model is correct, using either:
      - The BINOM.DIST (x, n, p, false) formula in Excel, or
      - The probability mass function for the binomial distribution:

$$P(x|n,p) = C_n^x p^x (1-p)^{n-x}$$

where:
x is the number of exceptions,
p is the probability of an exception for a given VaR confidence level,
n is the number of trading days in the sample, and
$C_n^x$ is the binomial coefficient with value $\frac{n!}{x!(n-x)!}$ (= COMBIN (n, x) in Excel).

   b) Compute the probability of observing k or more hits, given a correct model, as the sum of the outputs from step 2a) for all x between 0 and k-1:

$$1-\sum_{x=0}^{k-1} P(x|n,p).$$

   Alternatively, steps 2a) and 2b) can conveniently be summarised by the cumulative version of the Excel function. The probability of observing k or more hits is then [1–BINOM.DIST(k-1, n, p, true)].

3) Repeat steps 2a) and 2b) until the probability of observing k or more hits is below, or just slightly above, the significance level chosen under step 1). Then k-1 qualifies as the highest acceptable number of hits.

4) Count the number of hits in the sample. If this number exceeds the highest acceptable number defined under 3), reject the null hypothesis and conclude that the model provides inadequate unconditional coverage. Otherwise do not reject the null hypothesis.

## Annex 2

### Two-tailed test of unconditional coverage – Kupiec's test

Given N observations and n hits, our null hypothesis is n/N= α. The test statistic is computed as:

$$LR_{POF} = -2\ln\left(\frac{\alpha^n(1-\alpha)^{(N-n)}}{\left(\frac{n}{N}\right)^n\left(1-\frac{n}{N}\right)^{(N-n)}}\right)$$

If the VaR model is accurate, and given a sufficiently large sample, LRPOF is distributed like a $X^2(1)$([12]).

Therefore, using the critical values from the $x^2(1)$ distribution, we reject the null hypothesis:

- At the 10% significance level if $-2\ln(LR) \geq 2.706$,
- At the 5% significance level if $-2\ln(LR) \geq 3.841$, and
- At the 1% significance level if $-2\ln(LR) \geq 6.635$.

[12] '$X^2(1)$' reads Chi-Square distribution with one degree of freedom



## Annex 3

### Pearson's Q test of unconditional coverage at multiple levels

Given the number of hits that have occurred within each bin, the test statistic is calculated as follows:

$$Q = \sum_{i=1}^{k} \frac{\left(N_{(l_i,u_i)} - N(u_i - l_i)\right)^2}{N(u_i - l_i)}$$

Where:
$N(l_i,u_i)$ is the number of hits within sub-interval $(l_i,u_i)$,
N is the total number of observations, and
$(u_i - l_i)$is the width of the sub-interval.

Under the null hypothesis that the VaR model is accurate, Q is distributed like a $\chi^2(k-1)$ where k is the number of bins used in the test[13].

## Annex 4

### Christoffersen's independence test (1998 – consecutive violations only)

The test statistic is computed as:

$$LR = -2\ln\left(\frac{(1-\pi)^{(N_1+N_2)}\,\pi^{(N_3+N_4)}}{(1-\pi_0)^{N_1}\pi_0^{N_3}(1-\pi_1)^{N_2}\pi_1^{N_4}}\right)$$

Where:
- $N_1$ = number of non-violations on the days following a non-violation
- $N_2$ = number of non-violations on the days following a violation
- $N_3$ = number of violations on the days following a non-violation
- $N_4$ = number of violations on the days following a violation
- $\pi_0 = N_3/(N_1+N_3)$
- $\pi_1 = N_4/(N_2+N_4)$
- $\pi = (N_3+N_4)/(N_1+N_2+N_3+N_4)$ (hit ratio)

This is approximately centrally chi-squared with one degree of freedom. Therefore, using the critical values from the $X^2(1)$ distribution, we can reject the null hypothesis (that the probability of violation is the same whether the previous day was a violation or not) using the same critical values as presented in Annex 2 for Kupiec's POF test.

[13] To wit, '$x^2(k-1)$' reads Chi-Square distribution with k-1 degrees of freedom. This is how the sum of the squares of k-1 independent standard random variables is distributed.

## Annex 5

### Joint test of unconditional coverage and independence

Since two distinct equalities are being tested here, the test statistic (the sum of the test statistics for Kupiec's POF test and Christoffersen's 1998 test of independence) follows a chi-square distribution with not one but two degrees of freedom.

Accordingly, the null hypothesis will be rejected:

- At the 10% significance level if $-2ln(LR) \geq 4.605$,
- At the 5% significance level if $-2ln(LR) \geq 5.991$, and
- At the 1% significance level if $-2ln(LR) \geq 9.210$.

Note how these critical values are higher than the corresponding values for Kupiec's POF test, which are also used for Christoffersen's Markov test (see annex 2 and annex 4). In practice, this leads to a lower rate of rejection even for VaR models that generate more hits, or more consecutive hits, than a correct model. In other words, joint testing comes with a higher likelihood of a type 2 error.

Risk managers and conducting persons in charge of risk management may use the following checklist to (i) verify compliance with regulatory requirements, (ii) check they have set-up adequate ongoing testing.

| Regulatory Requirements | | Other Industry Practices |
|---|---|---|
| VaR Model Specifications | Backtesting Programe | |
| According to Box 15 of CESR 10-788, VaR needs to be computed with:<br><br>• One-tailed **confidence interval of 99%.**<br>• **Holding period equivalent to 1 month** (20 business days).<br>• Effective **observation period** (history) of risk factors of **at least 1 year** (250 business days) unless a shorter observation period is justified by a significant increase in price volatility (for instance extreme market conditions).<br>• **Quarterly data set updates**, or more frequent when market prices are subject to material changes.<br>• At least **daily calculation.**<br><br>A confidence interval and/or a holding period differing from the default parameters may be used provided the confidence interval is not below 95% and the holding period does not exceed 1 month (20 days). | According to Box 18 of CESR 10-788, the backtesting programme monitors the accuracy and performance of VaR model (i.e. prediction capacity of risk estimates). Main building blocks are:<br><br>• **Compare** for each business day the one-day **value-at-risk** measure generated by the model (VaR) for the end-of-day positions to the one-day **change of the portfolio value** (P&L) by the end of the subsequent business day.<br>• **Determine and monitor 'overshootings'** i.e. cases where the P&L exceeds the related one-day VaR.<br>• Back testing should **be carried out at least on a monthly** basis, subject to always performing retroactively the comparison for each business day.<br><br>If the number of 'overshootings' for the most recent 250 business days exceeds 4 in the case of a 99% confidence interval, additional analysis is warranted to identify sources of 'overshootings' and what measures (if any) are required to improve the accuracy of the model.In its 2014 activity report, CSSF has provided information regarding their review of VaR backtesting programmes. CSSF stated that the requirement described above should be supplemented with additional analyses such as:<br><br>• Assessing the VaR at different **confidence intervals.**<br>• Testing for of **clusters** of exceptions.<br>• Accounting for the **magnitude** of exceptions.<br>• Identifying **insufficient number of exceptions**. | Risk management needs to list and specify procedures for:<br>• **Statistical tests** to be performed monthly e.g.:<br>- Unconditional coverage (UC) i.e. that model has the expected coverage (99% in the standard approach).<br>- Independence (IND) i.e. that the occurrence of one exception doesn't affect the probability of occurrence of other exceptions (exceptions don't cluster together).<br>- Other tests e.g. joint test (UC+IND), size or timing of 'overshootings', etc.<br><br>• Monthly analysis **beyond statistical tests:**<br>- Review moves of the underlying risk factors.<br>- Check for large changes in positions.<br>- Probe for event risk that might not have been captured (FED tapering, Lehman bankruptcy, Fukushima meltdown, etc.).<br>- Contrast backtesting results to expected shortfall and stress-test results to gauge result plausibility.<br><br>• Frequency/triggers, for further **ad-hoc analysis:**<br>- Examine model's underlying assumptions.<br>- Graphical checks to capture forms of clustering not tested statistically.<br>- Pay attention to what happens in the tail.<br>- Contrast with backtesting results at the industry level. |

luxembourg fund industry

The Association of the Luxembourg Fund Industry (ALFI), the representative body for the Luxembourg investment fund community, was founded in 1988. Today it represents more than 1 300 Luxembourg-domiciled investment funds, asset management companies and a wide variety of service providers including depositary banks, fund administrators, transfer agents, distributors, law firms, consultants, tax advisers, auditors and accountants, specialist IT providers and communications agencies.

Luxembourg is the largest fund domicile in Europe and its investment fund industry is a worldwide leader in cross-border fund distribution. Luxembourg-domiciled investment structures are distributed in more than 70 countries around the globe, with a particular focus on Europe, Asia, Latin America and the Middle East.

ALFI defines its mission as to "Lead industry efforts to make Luxembourg the most attractive international centre".

Its main objectives are to:

**Help members capitalise on industry trends**
ALFI's many technical committees and working groups constantly review and analyse developments worldwide, as well as legal and regulatory changes in Luxembourg, the EU and beyond, to identify threats and opportunities for the Luxembourg fund industry.

**Shape regulation**
An up-to-date, innovative legal and fiscal environment is critical to defend and improve Luxembourg's competitive position as a centre for the domiciliation, administration and distribution of investment funds. Strong relationships with regulatory authorities, the government and the legislative body enable ALFI to make an effective contribution to decision-making through relevant input for changes to the regulatory framework, implementation of European directives and regulation of new products or services.

**Foster dedication to professional standards, integrity and quality**
Investor trust is essential for success in collective investment services and ALFI thus does all it can to promote high professional standards, quality products and services, and integrity. Action in this area includes organising training at all levels, defining codes of conduct, transparency and good corporate governance, and supporting initiatives to combat money laundering.

**Promote the Luxembourg investment fund industry**
ALFI actively promotes the Luxembourg investment fund industry, its products and its services. It represents the sector in financial and in economic missions organised by the Luxembourg government around the world and takes an active part in meetings of the global fund industry.

ALFI is an active member of the European Fund and Asset Management Association, of the European Federation for Retirement and of the International Investment Funds Association.

To keep up to date with all the news from the association and the fund industry in Luxembourg, join us on LinkedIn (The Luxembourg Fund Industry Group by ALFI), Twitter (@ALFIfunds), Youtube, Vimeo or visit our website at www.alfi.lu.

*March 2017*



**ABC of VaR Model Backtesting**

alfi

guidelines