International Journal of Business and Management Study – IJBMS
Volume 2 : Issue 1 [ISSN : 2372-3955]
Publication Date: 30 April, 2015

Probability and Pertinence

Adela Baho, Alain Ruttiens

***Abstract*— In our today culture, the availability of a quantitative measure is highly comforting, without to speak about the opportunity that such a quantitative measure represents within the framework of developing quantitative, descriptive models, in the financial, economic or any other field. However, one should be wary about the comfort of a quantitative measure, which may be fallacious in some cases, to the extent that it could seriously affect some decision/management process. In this paper, we are concerned with a specific subset of quantitative measures, namely, probability measures, in the particular case of (very) low probabilities. This paper is about the *pertinence* of such a (very) low probability measure: to what extent is this measure pertinent, with respect to its use, in financial, or, more generally, economic decisions? We propose a quantitative measure of this degree of pertinence, and apply it to an example of Value-at-Risk calculation.**

I. The validity of a (very) low probability calculation

In our today culture, the availability of a quantitative measure is highly comforting, without to speak about the opportunity that such a quantitative measure represents within the framework of developing quantitative, descriptive models, in the financial, economic or any other field. However, one should be wary about the comfort of a quantitative measure, which may be fallacious in some cases, to the extent that it could seriously affect some decision/management process. In this paper, we are concerned with a specific subset of quantitative measures, namely, probability measures, in the particular case of (very) low probabilities, called “tails probabilities” hereafter.

The use of tail probability measures may be problematic at two levels. The first level is the validity of a probability calculation, and has been extensively studied. This paper is rather about a second level, that is, the *pertinence* of such a measure: to what extent is this measure pertinent, with respect to its use, in financial, or, more generally, economic decisions?

Adela Baho

Aix-Marseille School of Economics, Visiting Lecturer

France

Alain Ruttiens

ESCP Europe, Paris, Affiliated Professor

France

Let us first compare two extreme cases of probability calculations,

- The case of a lottery with 1 million tickets and 1 single prize: the probability to win is 10^{-6};
- The case of the incidence of a major earthquake (M ≥ 7.0) in the Paris area within the next 50 years: the probability of such an incidence has been computed as 0.02 % before 2050[1];

In the first case, the probability determination is obviously 100% accurate. In the second case, the earthquake probability of 0.0002 is valid to the extent of the accuracy of its calculation methodology, and subject to a sampling size issue (i.e., the set of relevant historical data).

One can find numerous studies about the errors which may affect the determination of a probability measure, in the particular field of time series of financial markets data. The main problems arising when computing probabilities and related functionals such as their corresponding distributions – through their moments – are:

- The quality of the input data, in particular, time series of data: in other words, should the presence of outliers be considered, or not, as erroneous data (not rectified by the data provider)? The problem is all the more crucial if we are interested in tails probabilities: to what extent an extreme data (for example, a financial return), that could be taken into account in a tail probability measure, must be considered as a reliable data, even as an outlier? For more about outliers detection, cf. for example Rousseeuw, Leroy (2003);
- The quantity of available input data, which will directly affect the precision, and therefore the confidence interval of the measure: in many instances we do not dispose of long enough series of data;
- The frequency of the data: higher frequency data allow easier for huge enough data sets, but, for example in financial time series, erroneous data have less chances to have been corrected by the data provider on a higher frequency base;
- The sampling procedure used to compute the probabilities. Market data are discrete by essence: using them for computing continuous functionals poses the challenge of determining if the discontinuities between successive data are resulting from the discreteness of the sampling, or from actual jump dynamics. Indeed, even by using the full set of available, discrete, data, they are nevertheless making up a sample vis-à-vis the continuous series that will correspond to some continuous functional. This problem has been investigated by Aît-Sahalia (2002);

[1] Source: TX Earthquake report, Texas (USA).

- The hypotheses that can be postulated about a type of probability distribution (Gaussian, or others): they obviously affect more particularly the shape of its tails;
- The degree of homogeneity of the used data over time, with respect to their probability distribution, namely, the degree of stationarity of the distribution. Actual financial time series do actually present some lack of stationarity (cf. Taylor, 2007), and unfortunately,
  - the longer the time series, the less stationary,
  - the higher the moment of a distribution, the higher its variability over time.

Most of these error sources are all the more affecting rare events, quantified by (very) low probability levels. For example, the need for a sufficient size of the available data set, combined with enough stationarity. Or the difficulty of assigning an adequate probability distribution to observed tails probabilities.

In the following sections of this paper, we will not take into consideration to what extent the probability measures are affected by above problems.

II. The pertinence of a (very) low probability measure

Excluding the particular case of computing an exact probability form a sample equal to the full population (not a realistic situation in the business area), even by considering probability measures are resulting from a precise and reliable enough methodology, the next question is thus: Is it reasonable to include a probability measure in a calculation that must lead to a decision, or that will affect a decision? To be affirmative, the answer to this question must necessarily involve a sufficient degree of pertinence of the measure involved in the calculation, in other words, that this measure satisfactorily represents the reality or the reliability of the phenomenon which has been quantified in the measure. If it is not the case, it makes sense that the use of such a probability measure would better be restricted to a stylized fact, that is, to its actual equivalent of a qualitative assessment (from "very low probable" to "highly probable"). So that the previous question can be reformulated as: To what extent a probability measure should be valid enough to be safely used in a quantitative way, within the framework of a decision process? And the answer is: It depends on its degree of pertinence.

A. *A qualitative approach of the pertinence*

Intuitively, at first sight a (very) low probability should be considered as less pertinent than a (very) high probability: for example, it should make sense to use a (very) high probability measure in computing the size of a potential profit or loss. On the other hand, if one has to consider the building of a plant in the area of Paris, one may probably consider the risk of a major earthquake. If the probability of such an earthquake is of 0.02 % (cf. section 1), one will presumably not involve this data in the feasibility study, even if, qualitatively speaking, the low probability of such a major issue could make part of the study.

But a "degree of pertinence" should better be quantified, in order to handle it in a more precise way:

B. *A tentative, quantitative measure of the pertinence*

Let us denote p a probability measure and π its pertinence, looking for a $\pi=f(p)$ relationship. A trivial solution would consist in posing $\pi=p$, translating in the simplest way the fact that the lower the probability, the lower its pertinence. But such a proposition looks too simplistic, since there is no reason to admit that only the certainty ($p=1$) would be fully pertinent, nor that the degree of pertinence would not differ from the probability level itself.

The proposed, more grounded, pertinence measure rests upon the *confidence interval* associated to a probability measure. For a probability measure p, the confidence interval of an observed value p of this probability, associated to an error α, is given by

$$P\left(\bar{p}-\sqrt{\frac{pq}{n}}\times z_{\alpha/2}\le p\le \bar{p}+\sqrt{\frac{pq}{n}}\times z_{\alpha/2}\right)=1-\alpha \qquad (1)$$

, where $q = 1 – p$, n is the size of the sample used to compute p, and $z_{\alpha/2}$ is the z-score[2] corresponding to $\alpha/2$.

For a given p, reducing the width of the confidence interval to a suitable error level α implies to compute p with a sufficient sample size n.

If we want that n be so that the error on p would be for example +/- $\alpha/2$ % of p, this implies

$$\bar{p}+\sqrt{\frac{pq}{n}}z_{\alpha/2}=\bar{p}(1+\alpha/2)$$

, that is a sample size of

$$n=\frac{pq}{\left(\frac{\bar{p}\ \alpha/2}{z_{\alpha/2}}\right)^2} \qquad (1bis)$$

We can also compute the width of the "relative confidence interval" w_p , expressed in percent of p as:

[2] Here, the distribution of p is implied to be Gaussian, but this will not affect our further reasoning.

$$w_p = 2 \times \sqrt{\frac{pq}{n}} \times z_{\alpha/2} \times \frac{100}{p} \qquad (2)$$

This formula is usually applied on an "ex-post" basis to determine the confidence interval, for a given error and sample size. But we can view it on an "ex ante" basis as well: once the probability of a certain event has been computed with some error margin, in order to observe the occurrence of the event with this same probability, one must consider it with the same error margin, and on a sample of the same size (*n*) of future data. But a smaller ex post sample size will not reflect this probability of occurrence.

To illustrate this reasoning, let us consider the case of an event which has been determined to occur with a probability of 10%.To ensure that this probability is valid with a error margin of + or - 1% (corresponding to a z-score of 2.3263487), such as $0.09 \leq p \leq 0.11$, (1bis) says that it should have needed a sample of n = 487,070 (supposing the distribution of is Gaussian). Let us further focus on the case of probabilities of events to be observed in the course of the time, that is, the case of time series of variables, more specifically, on a daily basis. So that, to actually observe in the future the occurrence of this event with a probability of p= 10% within such a + or - 1% error margin, we need to cover a future period of n= 487,070 days. The shorter the n, or future period of time, the broader the confidence interval of p will be. At a rather standard horizon of time of 100 days or within n=100 trials, and for a given error margin of + or - 1%, hence a given value of $z_{\alpha/2}$ of 2.3263487, (2) gives 140% (rounded) as the width of the relative confidence interval corresponding to p= 10%, and from (1), one computes that during these 100 days, the event will actually occur with a probability range of 140% of 10% around p=10%, i.e.,10% + or – 7%, that is, comprised between 3% and 17%! As a result, a decision based on the premise that such an event would occur with a 10% probability during the next 100 days is at least questionable.

Coming back to the general case, we can build a pertinence measure that would reflect the sample size needed for reaching a given error level in the occurrence of an event. To scale such a measure, we propose, quite arbitrarily but realistically in many actual decision-making problems, to scale the measure in two ways:

- By considering a ≥ 50% probability as fully pertinent (π=1);
- For two different probability levels, by considering their relative confidence interval with respect to an identical sample size of n for the occurrence of the event.

Based on this scaling, for a given probability p_i, denoting $p_{0.5}$ as the reference level for a $\pi=1$, we posit

$$\pi(p_i) = \frac{\textit{relative confidence interval of } p_{0.5}}{\textit{relative confidence interval of } p_i} \qquad (3)$$

The rationale of this pertinence measure is the following: if one admits, as a reference point, that a probability of 50%, characterized by its relative confidence interval $w_{0.5}$, can reasonably considered as fully pertinent (although to be verified within some small enough error margin), the relative confidence interval w_{pi} for a probability p_i < 50% being larger, the ratio of both intervals will be reduced accordingly. Moreover, the ratio (3) presents the advantage of being independent of both α and n.

Applying relationships (2) and (3), we obtain the table and the graph below. The graph also shows the comparison between the pertinence measure and a trivial determination of pertinence ($\pi = 2p$, assuming π = 1 for $p \leq 0.5$).

Intuitively, the pertinence measure looks realistic: first, with decreasing probabilities, and down to a probability level of about 15%, π decrases in an approximately linear way, but at a slower pace than the trivial solution; but below this level of about 15% probability, π drops goes at a higher speed towards 0. This would imply that the taking into account of a (very) low probability measure into a quantitative calculation with respect to some decision process becomes more and more questionable.

Incidentally, it should be noted that the proposed pertinence measure is based on a Gaussian distribution through the above calculation of confidence interval. However, this confidence interval could significantly differ if said distribution is not Gaussian. In many fields, such as quantitative finance, the actual determination of the parameters of a non-Gaussian distribution is difficult, given the sensitivity of these parameters, subject to changes over time (non-stationarity). This should suffice to favor a robust pertinence measure that could be considered as "neutral", that is, not being based on any other (non-Gaussian) distribution.

III. Potential applications of the pertinence measure π of a (very) low probability

The use of the pertinence concept, and of its proposed measure, can be considered in a broad range of applications, related to any kind of decision making processes, such as, for example, within the framework of business plans, or in some particular cases of accounting principles. Or also in the much sensitive medical field, where statistics play a key role (pharmaceutical research, diagnosis): as an example, in a recent interview[3], professor Didier Sicard, past president of the French *Comité Consultatif National d'Ethique*, declared "To know that you have 2.1% chances more than your neighbor to

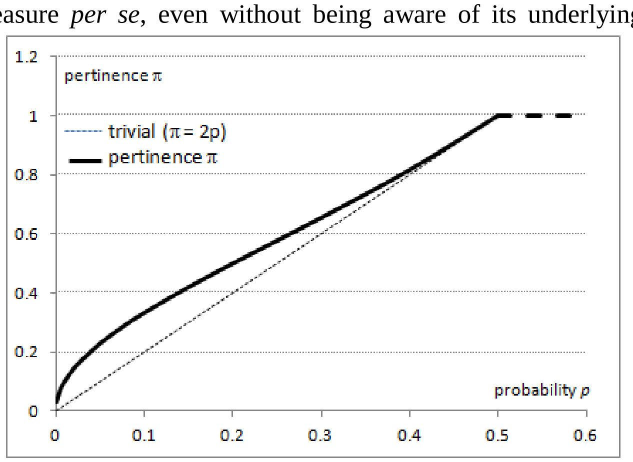
[3] In LE POINT, 17 October 2013

probability	relative Confidence interval *(in % of probability)*	pertinence	probability range (for a = 1%) *min:*	*max:*
0.001	1471	**0.03**	0[a]	0.01
0.005	656	**0.07**	0	0.02
0.01	463	**0.10**	0	0.03
0.02	326	**0.14**	0	0.05
0.025	291	**0.16**	0	0.06
0.05	203	**0.23**	0	0.10
0.075	163	**0.28**	0.01	0.14
0.1	140	**0.33**	0.03	0.17
0.15	111	**0.42**	0.07	0.23
0.2	93	**0.50**	0.11	0.29
0.25	81	**0.58**	0.15	0.35
0.3	71	**0.65**	0.19	0.41
0.4	57	**0.82**	0.29	0.51
0.5	47	**1.00**	0.38	0.62
> 0.5		**1.00**		

a. corrected to 0 when the calculation leads to a negative value)

develop such or such disease is a derisory information" (translated from French), showing an intuitive perception of our pertinence issue.

Interestingly, the lack of pertinence showed here above when considering (very) low probabilities, fits with a key aspect of the Prospect Theory. As Richard Zeckhauser (2010) writes, "Whether drawing from Prospect Theory of observation, it seems clear that individuals draw insufficient distinctions among small probabilities". The author adds: "Some strong supporters of behavioral decision theory, however, think it is our norms that are misguided, and that the way the brain naturally perceives outcomes, not the prescriptions of decision theorists and economists, should be the guidelines ". To some extent, we could appreciate the human wisdom, consisting in not granting such an importance to a small probability measure *per se*, even without being aware of its underlying pertinence metrics. This of course does not mean just to ignore a small probability, but, somewhat wisely, to take it into account on a qualitative rather than on a quantitative basis.



IV. Example of application in the quantitative finance area

In this section, we will restrict our applications to the field of quantitative finance, and more particularly with respect to time series data.

A. *Financial risk management and the use of a VaR measure*

Let us consider here the application of the pertinence concept and measure, to a growing concern in the market finance area, namely, the Value-at-Risk (VaR), which plays a key role in the Basel III regulations for banks, and, more recently, as a risk management tool in funds, also in UCITS funds.

For example, let us consider a fund invested for $ 100 million, which a long enough time series of past performances presents an average return μ of 10 % and a standard deviation σ of 15%, on an annual basis. For a (99%, 100 days) VaR, the left hand tail of the returns distribution goes from -∞ to:

$$\mu - 2.3263\sigma = 0.10 - 2.3263 \times 0.15 = -0.2489$$

, so that its (99%, 100 days) VaR is[4]

$$\$\ 100\ million \times 0.2489 \times \sqrt{\frac{100}{250}} = \$15.74\ million$$

In absence of any consideration about the degree of pertinence of the probability level, a 99%-VaR means that there is 1% of chances – here, 1 day on 100 days – that the fund could support a loss estimated to $15.74 million (needless to add that the VaR measure does not represent the maximum possible loss).

Now, applying (1bis), we can determine that this 1% probability measure would need a series of more than 5.3 millions of days in the future, in order to be actually observed with a precision degree of + or -1% (of this 1% level)... Viewed in another way, applying the previous relationships, a loss of $15.74 million over a horizon of the 100 next days only, may actually occur with a probability ranging from 0% chances up to 3.31 % (i.e. up to more than 3 times more frequently), within our relative error margin of + or – 1%. Keeping this into account, we would rather have to say that an

[4] Assuming a Gaussian distribution of returns, with μ and σ constant over the next 100 days, and a year counting 250 business days.

estimated loss of $15.74 million has actually a probability of up to 3.31% instead of strictly 1% chances to occur within the next 100 days to come.

Said in other words, the 1% VaR measure is far from pertinent, quantitatively speaking. And based on our pertinence scale, we could value this pertinence as π=10% only (cf. above table). Instead, using a 5% VaR measure would somewhat improve its pertinence, with a π of 23 %, i.e. more than twice.

In this example, the accuracy of the VaR at a 1 % level may be viewed as rather strict, namely of 1 % + or -1% of 1%, or a VaR between 0.99% and 1.01%. Given such a small variation, we can suppose that the loss amount of $15.74 million is very precise: the exact amount resulting from the above calculation is indeed $ 15,741,818. So that, by linearizing the distribution around the 1% level, with the said precision level the loss amount would vary between $ 15,584,400 and $ 15,899,236. Economically speaking this amount is not significant. But what if we loosen the VaR level? With a precision of 10 % in absolute value of this 1% level, that is, a VaR level between 0.9% and 1.10% - which does not affect the pertinence measure, – would still require (by using same formulae as above) about 1,6 million days to be verified in the future. And the loss of $15.74 million over a horizon of 100 days only, may actually occur with a probability ranging from 0% cup to 2.28 %, which is still not compatible with the spirit and the objective of the VaR as a risk measure.

One could even further lower the accuracy around the 1% VaR measure, leading to a shorter time horizon to actually verify this probability level, as well as to a narrower probability range of observing the loss of $ 15.74 million. But at the same time, the precision on the loss amount will further deteriorate. As an example, with a 1% VaR ranging from 0.8% to 1.2%, it would still need about 700,000 days to be verified, and the probability of occurrence of the loss during the next 100 days is still between 0.16% and 1.84%.

Conversely, by using the same formulae, we could compute the accuracy of the 1% VaR level which would be verified in the next 100 days. The calculation gives a so-called 1% VaR level ranging from 0.5039 to 1.4961%. In this case, during the next 100 days, the probability of loss would shrink between 0.99% and 1% and the loss amount would be seriously affected. Furthermore, assuming, for sake of simplicity, a linear dependence of the loss amount around the 0.5039 to 1.4961% probability range, the theoretical loss of $ 15.74 million would actually range from 7.93 million and 23.55 million, which is no more acceptable with respect to the VaR as a risk measure.

Further to the VaR itself, regulators require an ex-post backtesting of the VaR, to check to what extent its calculation is grounded. This backtesting is usually done by comparing the actual occurrence of losses exceeding the VaR level to the VaR probability level, over a period of one year, or about 250 business days. This is usually performed by using the Kupiek and/or Christoffersen tests. Unfortunately, by doing so, the lack of pertinence of the VaR 1% probability level is exacerbated because the comparison with actual exceeding losses is counted on a too small data or time interval, such as 250 data or 250 days, that is, far below a satisfactory confidence interval.

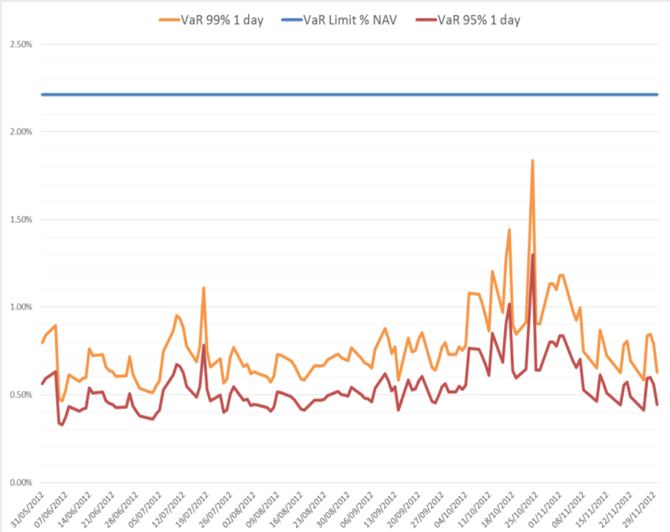
Let us further illustrate this with another real world example from the risk management of a UCITS fund. A UCITS fund can be managed according to one of the two available risk measures: the commitment approach (sum of notionals adjusted by netting and/or hedging) or the Value-at-Risk approach. In our example, the fund has adopted the VaR approach. In such a case, the regulator asks for an annual backtest in order to validate the VaR model.

Our € 260 million fund is invested in various asset classes making use of derivative instruments. It has, according to the prospectus, a VaR[5] limit or budget of 2.21% of the NAV of the fund. That is, 1 day out of 100, the expected loss of the fund should not be higher than 2.21% of its Net Asset Value or € 5.46 million. The graph below illustrates a time series of 6 months or 180 days during which the 99% VaR reached a level of 1.84% or € 4.78 million. Applying (1bis), we see that one would need more than 5.3 million observations in order for this probability measure to be in the interval of 0.993% and 1.027% around the 1% average level. But at the same time the degree of pertinence π is also low, namely at only 10%.

Now, according to our measure of pertinence, the decision-maker (fund manager) might decide to increase it to 23%, meaning a probability of 5%. In that case the VaR 95% would be lower as illustrated in the next “Backtesting” graph. However, this graph displays 6 breaches for the 95% VaR instead of only 2 breaches for the 99% VaR. The model with 95% VaR will not be accepted in that case by the regulator. Therefore, although the degree of pertinence is much higher in the second case, the decision-maker will chose the 99% VaR which gives rise to less breaches.

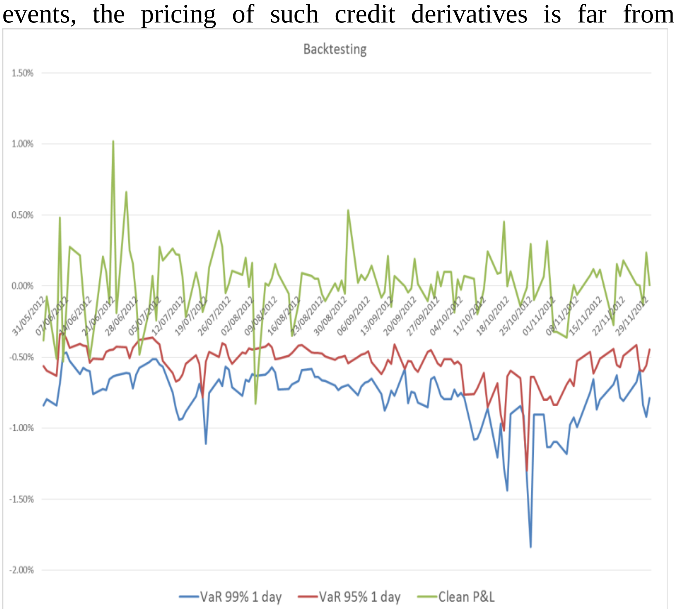
The decision-maker can even go further in this case. Given that the VaR budget is significantly high (20% higher than the highest VaR level) it can lower the limit of 2.21% of the NAV, let say to 2%. This decision is not without consequences on the risk-profile of the fund. The VaR budget is used for what is called SRRI (Synthetic Risk and Reward Indicator) which gives to the UCITS investor a description of the maximum market volatility risk of the fund. Lowering the limit means that the fund is prone to lower risk, which is misleading given that the VaR has not changed (under the assumption that the model is accepted).

[5] The VaR model used here is a mix of GARCH and EWMA models. The diagonal of the variance-covariance matrix is estimated via GARCH, and correlations via EWMA.



B. *Other applications in finance*

There are many other fields where (very) low probability measures are common in quantitative finance, such as for example in the pricing of (D)OTM options. A more sensitive topic concerns the valuation of a default or credit risk, and the related derivative products such as CDS (credit default swaps). By nature, a probability of default is a very low probability. Determining the "fair" (or theoretical) price of a CDS is necessarily involving some probability measure that is in the range of, say, 1 % or below. Default situations being rare events, the pricing of such credit derivatives is far from satisfactory[6]. Furthermore, involving a very low probability measure into such a calculation is questionable, given its lack of pertinence.



Finally, let us consider the case of portfolio optimization based on a utility function. Assuming some distribution of portfolio returns, with respect to the Sharpe-Markowitz portfolio theory, the utility of a portfolio is typically viewed as a function of both the returns and the variance (hence, volatility, risk) of the observed past returns of the assets making up the portfolio. Each of the various possible values of a portfolio can be associated to the probability of obtaining some utility value. So that the expected value of the utility function is a sum of different possible portfolio values, weighed by their probability of occurrence. In particular, some high portfolio values will be associated to (very) low probabilities. Clearly, such an approach is questionable, the moment we worry about the pertinence of these low probability levels.

V. Conclusion

In many decision-making processes, the use of tail probability measures, i.e., of (very) low probabilities, may be problematic at two levels. The first level is the validity of a (very) low probability calculation, given the difficulties to determine them; this aspect is not covered here. This paper is rather about a second level, that is, the *pertinence* of such a measure, and proposes an objective, quantitative measure, called "π", of the degree of pertinence of a probability metrics. The proposed determination of this pertinence π measure makes use of the confidence interval which is associated to any probability measure, and has been scaled so that a probability ≥ 50 % is conventionally equal to 1, and decreases down to 0 for probabilities < 50 %. So that the pertinence π, associated to a given probability percentage, reflects the precision of this percentage through its link to the confidence interval which is associated to. This allow thus to assess the degree of pertinence of a probability measure that would be involved in any kind of calculation leading to a management decision.

Altogether, before incorporating a probability measure into her decision process, a decision-maker would do well to take into account the degree of pertinence π of this probability, and probably give up to make use of this probability measure in her calculations if its pertinence π is judged too small, i.e., below some minimum threshold. Conversely, the soundness of a (financial) management decision involving some probability measure can be assessed by considering the pertinence π of the probability involved in. Alternatively, facing (very) low probabilities, instead of dealing with (very) low probability measures, characterized by a (too) low pertinence level, a decision-maker should rather incorporate it into her decision-making process on a qualitative rather than on a quantitative basis.

[6] See for example Alain Ruttiens, *Mathematics of the Financial Markets*, Wiley, 2013.

References

Y. AÏT-SAHALIA, "Telling from discrete data whether the underlying continuous-time model is a diffusion," Journal of Finance, vol. LVII No 5, 2002, pp.2075-2112.

P. J. ROUSSEEUW, A. M. LEROY, "Robust Regression & Outlier Detection," Wiley, 2003.

S.TAYLOR, "Modelling Financial Times Series," World Scientific Publishing, 2nd ed., 2007.

R. ZECKHAUSER, "Investing in the Unknown and Unknowable, in Ar. S. WOOD (Editor), Behavioral Finance and Investment Management," CFA Institute, 2010.