Computing the Value at Risk of a Portfolio: Academic literature and Practionners' response.

Olivier Guéant

Contents

Introduction

Value at Risk, often referred to as VaR, is one of the most commonly used risk measures in the financial industry. Value at Risk is indeed a very simple concept that gives for a portfolio and a given confidence level p, a threshold of loss (over a given time horizon) that is expected to be exceeded only $(1-p)\%$ of the time.

This concept of Value at Risk is often supplemented by other more complicated risk measures such as the Expected Shortfall[1] (ES), the Tail Value at Risk (TVaR) or more confidential indices such as the Entropic risk measure, because it is known to violate one of the hypotheses characterizing the so-called *coherent risk measures* – it is not sub-additive – and since it does not reflect the entire tail of the PnL distribution.

Notwithstanding its well-known weaknesses, VaR is central to risk measurement and risk management and often perceived as a figure easy to communicate on. Also, VaR plays a central role for regulatory reasons. VaR figures are indeed underlying the computation of banks' regulatory capital and the accuracy of Value at Risk estimates is therefore of utmost importance for both regulatory purposes and risk management.

But, if Value at Risk is a very simple concept – a quantile in the distribution of the PnL of a given portfolio –, its definition is by no means constructive. The theoretical distribution of the PnL of a portfolio is indeed not observable and must be estimated. Computing the Value at Risk of a portfolio is therefore a statistical problem but not only. Since portfolios under scrutiny are often large and composed of complex financial assets, estimating the distribution of a portfolio's PnL requires approximations and relies on asset pricing: a finite number of relevant risk factors must often be appropriately chosen and the portfolio components matched to these risk factors before being priced.

Being at the same time a matter of approximations and a statistical problem, computations of VaR figures can be carried out using various methods and these methods are usually divided into three groups:

- Historical simulations
- Parametric methods, also called analytical methods

[1] This risk measure is sometimes called Conditional Value at Risk (CVaR).

- Monte-Carlo methods

Historical simulations consist first in choosing a certain number of relevant risk factors depending on the portfolio, and then in *replaying* data from the past behavior of these risk factors to evaluate what would have been the evolution of the price of the portfolio under scrutiny and eventually figuring out the potential losses when holding it during a certain period. Simple historical simulations date back to the very introduction of Value at Risk but the methodologies have been improved by academics, notably to diminish the lag between an evident increase in risk and its contribution to Value at Risk estimates. These methodologies are described in Part 1 of this document (BRW method, Hull and White approach, Filtered Historical Simulations) and we also discuss in this part good estimates of the extreme quantiles of an empirical distribution. The contribution of Extreme Value Theory (EVT) and quantile regressions to quantile estimation are indeed discussed and it appears that the tools coming from Extreme Value Theory are not used by most practitioners (see Part 4) whereas it potentially provides better estimates for extreme quantiles.
Historical simulations are used in many institutions and the main advantage, from a theoretical point of view, is that it does not require any calibration regarding the interdependence structure (correlations for instance) of the risk factors. Also, as documented in Part 4 of this document, it is based on a methodology with few hypotheses and thus, VaR figures based on historical simulations are easy to communicate on.

Early methods to compute Value at Risk were not all based on historical simulations and RiskMetrics popularized parametric (or analytical) methods in the mid-90s. Parametric methods are very fast methods to compute VaR figures that rely on both an approximation of the portfolio and a strong assumption on the distribution of the risk factors' returns. Parametric methods are numerous and they all share the advantage that, if not always given in closed-form, resulting VaR figures are easy to compute numerically. Most of these methods are based on an approximation of the portfolio using a Taylor expansion and thus rely on the Greeks of the assets in the portfolios. A consequence is that parametric methods do not need full repricing. However, even though the initial and well-known Delta approach has been supplemented by other approaches to take non-linearities into account, not all non-linearities can be handled and we present in Part 2 a large number of approaches that can be used to compute the Value

at Risk for vanilla portfolios. Parametric methods are used in banks – often to complement another VaR measure – and all practitioners seem to use gaussian risk factors.
In this second section, we also present some generalizations to non-gaussian risk factors of the usual parametric approaches.
Finally, parametric methods are presented that contribute to variance reduction techniques for Monte-Carlo simulations and Part 2 of this document is therefore linked to Part 3 on Monte-Carlo approaches.

Monte-Carlo simulations consist in the same type of simulations as in the historical approaches apart from the fact that the samples are not bounded to be based on past realizations of the risk factors. In fact, Monte-Carlo simulations rely on a calibrated distribution for the risk factors and draw scenarios from this joint distribution. The main advantage of Monte-Carlo simulations is that any distribution can be chosen for the risk factors but the counterpart is that a lot of parameters have to be estimated. Also, as for the historical approaches, Monte-Carlo approaches to VaR computations rely on full repricing and are time-consuming. Since indeed, Monte-Carlo simulations are known to converge very slowly, the number of draws must be large and we present variance reduction techniques that fasten the convergence. These variance reduction techniques, such as importance sampling or stratified sampling, seem to be rarely used in the industry for VaR computations whereas the time-consuming issue of Monte-Carlo simulations is really important (VaR figures must indeed be produced on a daily basis). In addition to the classical Monte-Carlo approaches, we also present a rather new approach in Part 3 based on stochastic algorithms. This approach may be seen as an alternative to Monte-Carlo with similar properties for practical use.

In the first three parts, we review the methodologies to compute Value at Risk and present some very recent approaches. However, in addition to the academic literature on the subject, we decided to carry out interviews of Chief Risk Officers and their teams in a panel of banks and insurance companies in order to discuss their models and exchange views on the academic literature. We also met teams from the French regulatory authority not to miss part of the picture on VaR computations since Value at Risk models are regularly audited by the Autorité de Contrôle Prudentiel. People at AXA, Natixis, Crédit Agricole CIB, HSBC, Barclays Capital and Autorité de Contrôle Prudentiel kindly accepted to exchange with us on the topic and we

are grateful to these people and their teams. The fruits of our discussions are summarized in Part 4[2] that exposes practices in the financial industry and the reasons underlying the choice of one or another model. The respective viewpoints of academics and practitioners regarding future research are also presented.

Before starting the presentation of VaR computation techniques, we have to say that this document focuses on the mathematical aspects of the computations. It is noteworthy that, throughout this text, the important question of the relevant risk factors to consider is not tackled. The reason for this is that it is an issue with no general answer, the answers depending strongly on the considered portfolio. In addition to the choice of the risk factors, another question is, for most methods, whether to consider additive variations or multiplicative variations (returns) to describe the evolution of risk factors. This question is not tackled here but it is of great importance in practice and the interested reader may look at the documentation [86] of RiskMetrics 2006 for an interesting discussion on this topic. Also, we do not really discuss the issue of time horizon. Bankers are required to compute Value at Risk over 10 days and 10 days may be thought of as being 1 single period or 10 periods of one day. In the former case the data is often made of overlapping returns and it raises the question of statistics with autocorrelated data. In the latter case, a rescaling factor is often used, usually[3] $\sqrt{10}$*, and it is a broad approximation. These questions are not discussed in this text although practitioners are confronted by them. We did however focus, for a given set of risk factors, on the various mathematical methodologies to compute VaR figures. Moreover, we must say that most methods we present (except perhaps Monte-Carlo methods) are adapted to short time horizons and not well suited to Value at Risk over long periods of one year, as required in the insurance industry for instance.*

To continue to clarify what this document is or is not, we must also clarify that the Value at Risk under scrutiny is the Value at Risk linked to market risk. Credit risk, or more exactly default risk, is indeed not tackled in this document and the interested reader can read [24] to have an overall idea on this issue. However, to speak about the evolution of the regulation, the Stressed-VaR can be computed with the same tools as those presented in this document.

[2]The opinions presented in this part are ours, based on discussions with the practitioners. None of the practitioners from the panel is to be responsible for what is written in Part 4.

[3]That figure however supposes gaussian risk factors and linear portfolios.

Part I

Historical approaches

1 Historical simulations and recent improvements

Amongst the most straightforward methods to estimate Value at Risk, the historical methodology is certainly the first one that has been used among practitioners. Still today, historical simulations to compute Value at Risk play a major role in banks and other financial institutions.
Basic historical simulations were initially used but many methods were introduced to limit the impact of the drawbacks linked to the most simple historical simulations. In what follows, we present the main characteristics of the historical approach and we detail the different improvement successively introduced in the literature: the BRW method [13], the Hull and White contribution [50] and the filtered historical simulations (also known as FHS) [8].

Extreme value theory and other techniques also improved the usual Value at Risk estimates and we dedicate the next two sections to quantile estimation techniques.

1.1 Introduction to historical simulations

The historical approaches to compute Value at Risk are non-parametric methods that broadly consist in *replaying* data from the past behavior of some risk factors and reevaluating the portfolio under scrutiny to figure out the potential losses when holding it during a certain period.

In its basic definition, it consists of three steps. First, we have to choose a certain number of risk factors $X^1, \ldots, X^n$ and we have to be able to price the different assets in the portfolio (and hence the portfolio) as a function of the different risk factors[4].
Once this first step has been carried out, the second step, which is the core of the basic historical approach, is to consider the empirical distribution of the losses the portfolio would have incurred

[4]This first step may include approximations in order to map the financial assets to the different risk factors.

over 1 period of time[5] when considering the historical data over a predetermined number of days. In other words, if for notational convenience we avoid path-dependent products, we may write the hypothetical *historical*[6] time series $(\mathcal{P}_{s,s+1})_{s\in[t-T,t-1]}$ of the portfolio's P&L (the portfolio composition being the current one) as:

$$\mathcal{P}_{s,s+1} = V(X^1_{s+1}, \ldots, X^n_{s+1}) - V(X^1_s, \ldots, X^n_s)$$

and the *empirical* distribution of the P&L, or symmetrically the distribution of the losses, consists of these observations over a certain period in the past, *i.e.* $s \in [t-T, t-1]$ where t is the current time and T the length of the period. More precisely, in this second step of the most basic historical simulations we compute the empirical cumulative distribution G of $(\mathcal{P}_{s,s+1})_{s\in[t-T,t-1]}$:

$$G(z) = \frac{1}{T}\sum_{s=t-T}^{t-1} 1_{\mathcal{P}_{s,s+1}\leq z}$$

Once this second step has been carried out, the last step consists in estimating the quantile corresponding to the Value at Risk one wants to estimate. This can be done directly through the computation of the $(1-p)$-quantile of the empirical distribution G or using more sophisticated tools: bootstrap, Extreme Value Theory, ... (see below).

If the first step deserves some discussion about the choice of the risk factors, this discussion is out of the scope of this document because it depends strongly on the portfolio and is specific to each case (the interested reader may get some insights about the risk factors in the description we make of the initial RiskMetrics framework in the next part of this document).

Coming to the second step, the underlying hypothesis is that the changes in the value of the portfolio are identically distributed over the period. This hypothesis, which is central in the basic historical approach, is an issue since one of the most common stylized facts of market

[5]Throughout this text, the Value at Risk will always be considered over 1 period that we arbitrarily regard as 1 day. Although it may be possible in some (rare) cases to transform a 1-day Value at Risk into a 10-day Value at Risk by multiplying the 1-day Value at Risk figure by $\sqrt{10}$, we think this is in general a very bad approximation. The same reasoning done for a period of one day can in fact be applied over any – sufficiently small – number of days (using sometimes rolling windows) and we prefer not to use any scaling factor.
Hence, the reader should keep in mind that our period of time is arbitrary.

[6]It is not the historical time series of the P&L in the usual sense but it rather consists of the P&L values if the same portfolio had been held for the last T periods.

data is volatility clustering, at odds with the *i.i.d.* hypothesis. Another direct consequence of the *i.i.d.* hypothesis is that values from the far past are assumed to be equally relevant as recent data. The BRW (Boudoukh, Richardson, Whitelaw) method (see [13]) and the Hull and White proposal (see [50]) try to correct this drawback.

However, the advantage of such an approach is that it is non-parametric. In other words, for each period, the evolution of the risk factors does not rely on any distributional assumption: it naturally incorporates fat tails and a potentially complex interdependence between the risk factors. This non-parametric dimension of historical approaches is the main advantage pleading for them being used.

Coming now to the third step, due the very discrete nature of the observations, estimating a quantile in the tails is subject to important variance. If the empirical observations consist indeed of a year of daily observations, *i.e.* around 250 observations, the quantile corresponding to the estimation of the Value at Risk with confidence level 99% will be the second or the third largest losses (or a convex combination between them if one wants to interpolate). Subsequently, it rather depends on the realizations of the risk factors than on their probability distribution. The consequence is that the Value at Risk calculated with a quantile of the empirical distribution G will be highly unstable, especially when considering a Value at Risk with a high confidence level with only few available data. To avoid this instability and provide more accurate figures, Extreme Value Theory (EVT) can be used and we will come to it in the next section.

In general, and apart from the two issues presented above, one of the many criticisms against the historical simulation approach[7] is that, being highly non-parametric, it may not take into account an increase in risk if the recent realizations were indeed subject to more volatility but did not correspond to losses. Similarly, regime switching may not be accounted for early enough if the first realizations after the regime switched are positive profits. The basic underlying issue, which we call the asymmetry bias, is that realizations of the P&L above the most recent estimate of Value at Risk[8] usually do not decrease the Value at Risk in the basic

[7]The interested reader will find a review of some of the main problems with the use of historical simulation in [75].

[8]In fact above $-$VaR.

historical approach, although new observations, be they positive or negative, should provide information. The Hull and White contribution to historical simulations partially solve this issue. More generally, we will see that another approach, lying in between historical simulations and Monte-Carlo simulations, has also been developed that attempts to bring a solution. This approach is often referred to as Filtered Historical Simulation (FHS).

1.2 The BRW and Hull and White methods

The basic historical approach considers the historical time series of the risk factors and builds a hypothetical time series for the 1-period P&L:

$$\mathcal{P}_{s,s+1} = V(X^1_{s+1}, \ldots, X^n_{s+1}) - V(X^1_s, \ldots, X^n_s)$$

Then, the cumulative distribution function G built from this time series $(\mathcal{P}_{s,s+1})_{s\in[t-T,t-1]}$ is considered and the appropriate quantile is calculated based on G (either directly or using more sophisticated methods).

By definition, G is defined by:

$$G(z) = \frac{1}{T}\sum_{s=t-T}^{t-1} 1_{\mathcal{P}_{s,s+1}\leq z}$$

The underlying hypothesis is that the observations of the risk factors are independent realizations of the same underlying random variable. This *i.i.d.* hypothesis introduces a bias since, because for instance of volatility clustering, the last values of the risk factors are certainly more relevant for our Value at Risk estimate than their values a few month before. For this reason, Boudoukh, Richardson and Whitelaw introduced the so-called BRW method (see [13]) in which more weight is put on the last observations. In other words, and formalized in line with our presentation, a decreasing sequence of weights is introduced:

$$w_1 \geq w_2 \geq \ldots \geq w_T > 0, \qquad \sum_{i=1}^{T} w_i = 1$$

Then, another cumulative distribution function G_{BRW} is introduced that takes account of these

weights:

$$G_{BRW}(z) = \sum_{s=t-T}^{t-1} w_{t-s} 1_{\mathcal{P}_{s,s+1} \leq z}$$

Typically, the authors proposed an exponentially decreasing influence of the data:

$$w_i = \frac{1-\lambda}{1-\lambda^T} \lambda^{i-1}$$

and they applied their approach using $\lambda = 0.97$ and $\lambda = 0.99$.

If this method puts more weight on the recent values of the risk factor and may improve the accuracy of Value at Risk estimates in a volatility clustering context, the observations cannot be assumed to be independent. Also, it is of no help to correct the asymmetry bias discussed above.

Contrary to the BRW method, the Hull and White approach, developed in [50], proposes a possible solution to limit both the *i.i.d.* bias and the asymmetry bias. This approach consists in updating the level of volatility of past data to match its current level.
In some sense, this approach relies purely on the empirical data for the dependence between the different risk factors but slightly changes the marginal distributions of each risk factor whose returns are artificially scaled. More precisely, the variance of the risk factors' returns are scaled to mimic the current level of volatility, thus adapting the time series to current volatility conditions, without changing the skewness nor the kurtosis indicators.

Mathematically, it means that the times series $(\mathcal{P}_{s,s+1})_s$ is replaced by a rescaled one, depending on current volatility :

$$\mathcal{P}_{s,s+1} = V\left(X_s^1 + \frac{\sigma_{t+1}^1}{\sigma_{s+1}^1}(X_{s+1}^1 - X_s^1), \ldots, X_s^n + \frac{\sigma_{t+1}^n}{\sigma_{s+1}^n}(X_{s+1}^n - X_s^n)\right) - V(X_s^1, \ldots, X_s^n)$$

where the variables σ_s^i are volatility indicators for each risk factor that can be calculated using for instance an exponentially weighted moving average of the risk factors' squared return[9]:

[9]Most of the time, we assume that there is no drift. If the empirical data exhibit drifts, it is common to remove this drift before applying any Value at Risk calculation. Also, we assume here that volatility indicators are computed on returns and not on variations. The former is adapted to risk factors that are homogeneous to prices while the latter is suited for risk factors such as interest rates or credit spreads.

$$\left(\sigma_{s+1}^{i}\right)^2 = \lambda \left(\sigma_{s}^{i}\right)^2 + (1-\lambda)\left(\frac{X_s^1 - X_{s-1}^1}{X_{s-1}^1}\right)^2, \qquad s \in [t-T+1, t-1]$$

and they consider $\lambda = 0.94$ as in the original RiskMetrics approach (see below).

Once the rescaling of the P&L time series has been carried out, the methodology is the same and we compute the cumulative distribution function G_{HW}:

$$G_{HW}(z) = \frac{1}{T-1} \sum_{s=t-T+1}^{t-1} 1_{\mathcal{P}_{s,s+1} \leq z}$$

Then a Value at Risk estimate is calculated using an estimate of the appropriate quantile (directly through quantiles of G_{HW} or using Extreme Value Theory for instance – see below).

This Hull and White approach is a better solution than the BRW method to solve the *i.i.d.* bias since the innovation process $\frac{1}{\sigma_{s+1}^1}\frac{X_{s+1}^1 - X_s^1}{X_s^1}$ can better be assumed to be made of realizations of independent variables, the volatility clustering having been factored out.

1.3 From historical simulations to filtered historical simulations

Another way to use the Hull and White idea of past returns rescaling is to build from historical data a set of possible standardized returns. We can indeed build a collection $\mathcal{C}$ of standardized returns by:

$$\mathcal{C} = \left\{ \left(\frac{1}{\sigma_{s+1}^i} \frac{X_{s+1}^i - X_s^i}{X_s^i} \right)_i \middle| \, s \in [t-T+1, t-1] \right\}$$

where the volatility process is still updated by exponentially weighted moving average:

$$\left(\sigma_{s+1}^{i}\right)^2 = \lambda \left(\sigma_{s}^{i}\right)^2 + (1-\lambda)\left(\frac{X_s^1 - X_{s-1}^1}{X_{s-1}^1}\right)^2, \qquad s \in [t-T+1, t-1]$$

Then, we can, as we would do for a Monte Carlo simulation, consider the possible future values of the portfolio using the current estimate of volatility and standardized return for the risk factors from $\mathcal{C}$.

Phrased in another way, we introduce a collection $\mathcal{C}_{\mathcal{P}}$ of potential P&L figures for the portfolio by:

$$\left\{ V\left(X_t^1 + \frac{\sigma_{t+1}^1}{\sigma_{s+1}^1}(X_{s+1}^1 - X_s^1), \ldots, X_t^n + \frac{\sigma_{t+1}^n}{\sigma_{s+1}^n}(X_{s+1}^n - X_s^n) \right) - V(X_t^1, \ldots, X_t^n) \middle| t - T < s < t \right\}$$

Then, we can proceed as for the classical historical methodology and build a cumulative distribution function $\tilde{G}_{HW}$:

$$\tilde{G}_{HW}(z) = \frac{1}{T-1} \sum_{\mathcal{P} \in \mathcal{C}_{\mathcal{P}}} 1_{\mathcal{P} \leq z}$$

This approach, which is another Hull and White approach, is an example of Filtered Historical Simulation (FHS). It is considered a historical method for obvious reasons but it rather works as a non-parametric[10] Monte-Carlo simulation in which the standardized returns of the risk factors are drawn from $\mathcal{C}$ and multiplied by the current estimate of volatility.

Filtered historical simulations were introduced and developed by Barone-Adesi et al. [8, 7] and can be carried out over multi-day periods[11] drawing daily returns from a precomputed collection of standardized daily returns (assumed to be independent realizations of a random variables). Apart from this multi-day setting, the FHS approach also allows for complex processes for the volatility and they propose for instance a GARCH(1,1) process for the risk factors.
More generally, FHS can be used with many models for the returns (see [3], [23] or [73] for examples). Model parameters may be estimated using quasi-maximum likelihood estimates and then, once the model is calibrated, the empirical residuals can be calculated empirically to constitute a collection of potential values for the innovation process. The same Monte-Carlo-like approach as above can finally be used, drawing values of the innovation process from this collection.

As for the other historical approaches, the main issue is that the number of available data is rather limited. Overall, in most computations, only a few hundred points may be available and estimating the first percentile of a distribution on these data may seem arguable. The next

[10] More precisely, the distribution of the standardized return is non-parametric but the variance is described by a parametric process.
[11] It may be interesting when evaluating path-dependent assets.

section on Extreme Value Theory discusses this estimation issue.

2 Extreme Value Theory

In all the preceding historical simulations and filtered historical simulations, we ended up, at the end of the second step, with a cumulative distribution function G that was expected to be an approximation of the true cumulative distribution function of the P&L. The problem that was left for the third and final step is the computation of an estimate of the quantile at stake in the computation of the Value at Risk.

The naive approach consists, when the data used to build the empirical cumulative distribution function G are $(z_1, \ldots, z_T)$, in using the order statistics $z_{(\lfloor (1-p)T \rfloor)}$ and $z_{(\lfloor (1-p)T \rfloor + 1)}$ and building for instance a linear interpolation between them.

However, this method only takes into account a small part of the information contained in the cumulative distribution function G and the measure is quite volatile since we are looking for the $(1-p)$-quantile, p being close to 1.

To improve estimates, bootstrap approaches can be used and a Jackknife estimate[12] is for instance proposed in [61].

In this section, we are going to present another way to account for the information available in the empirical cumulative distribution G. A theory has indeed been developed to estimate extreme quantiles and is part of the wider Extreme Value Theory[13] (EVT).

2.1 Introduction to EVT

Extreme Value Theory is usually best known for the limit distribution of the maximum value of a sample as the size of the sample tends to infinity (see [35, 41]). An important result of Extreme Value Theory is indeed that, for *i.i.d.* samples $(Y_1, \ldots, Y_T)$ of the same random variable Y, there exist constants a_T and b_T such that:

[12] A simple Jackknife is not suited to quantile estimation. One has indeed to delete at least $\sqrt{T}$ elements from each sample.

[13] For an overview of the usefulness of EVT in Finance, the interested reader may refer to [29, 30].

$$\frac{\max(Y_1, \ldots, Y_T) - a_T}{b_T} \to_d F$$

where the distribution F is either a Fréchet, a Gumbel, or a Weibull distribution, or in terms of cumulative distribution function:

$$F_\xi(y) = \exp\left(-(1+\xi y)^{-\frac{1}{\xi}}\right)$$

for $y \geq -\frac{1}{\xi}$ if $\xi > 0$, for $y \leq -\frac{1}{\xi}$ if $\xi < 0$ and for $y \in \mathbb{R}$ if $\xi = 0$ – in that latter case $F(y) = \exp(-\exp(-y))$.

We say that Y is in the max-domain of attraction of the corresponding distribution F_ξ.

As above for the maximum, it is common in Extreme Value Theory to consider extremes in the right tail. This is a convention at odds with the issue of Value at Risk computation but we will stick to it since one just need to consider distribution of losses instead of P&L.

Here, we are not interested in the distribution of the maximum but rather in the distribution of a quantile in the tails. To this purpose, EVT provides an important asymptotic result for peaks over threshold. This result, known as Pickands-Balkema-De Hann's theorem says (see [4, 72]) that when Y is in the max-domain of attraction of F_ξ then for large thresholds u, the probability $\mathbb{P}(Y - u \leq y | Y > u)$ – which corresponding to the evaluation of the cumulative distribution function of the peaks above u – can be approximated by $H_{\xi,\sigma(u)}(y)$ where $H_{\xi,\sigma(u)}$ is the cumulative distribution function of a generalized Pareto distribution (GPD):

$$H_{\xi,\sigma(u)}(y) = 1 - \left(1 + \xi\frac{y}{\sigma(u)}\right)^{-\frac{1}{\xi}}$$

for $y \geq 0$ if $\xi \geq 0$, for $y \in [-\frac{1}{\xi}, 0]$ if $\xi < 0$ – if $\xi = 0$, $H_{0,\sigma(u)}(y) = 1 - \exp\left(-\frac{y}{\sigma(u)}\right)$.

The important point to notice is that only σ depends on u while ξ is inherited from the max-domain of attraction of Y.

2.2 Quantile estimation with EVT

Now, we want to estimate the quantile of a given distribution for which we know a few hundred realizations. To apply EVT we need to have independent observations. Although this is not the case in the basic historical approaches, the Hull and White method or the filtered historical simulations may be thought to provide *i.i.d.* realizations of the same variable.

The idea is that extreme quantiles (such as the first percentile, or here the 99^{th} one) cannot be estimated precisely but less extreme quantiles can. In other words, if we consider a 95% empirical quantile a good approximation of the true 95% quantile, we are going to use this value as a threshold and build an estimate of the 99% quantile using the distribution of the peaks over the 95% empirical quantile.

Formally, if one wants to estimate the α-quantile of a distribution for α close to 1, then one will rely on a less volatile estimate $Y_{(\lfloor \beta T \rfloor)}$ of the β-quantile for $\beta < \alpha$ and use the following decomposition:

$$\mathbb{P}(Y > y) = \mathbb{P}(Y > Y_{(\lfloor \beta T \rfloor)})\mathbb{P}(Y - Y_{(\lfloor \beta T \rfloor)} > y - Y_{(\lfloor \beta T \rfloor)} | Y > Y_{(\lfloor \beta T \rfloor)}), \qquad y > Y_{(\lfloor \beta T \rfloor)}$$

Hence, inverting this expression, we can approximate the tail function through:

$$\mathbb{P}(Y > y) = (1 - \beta) H_{\xi, \sigma(Y_{(\lfloor \beta T \rfloor)})}(y - Y_{(\lfloor \beta T \rfloor)})$$

Hence, an EVT estimate of the α-quantile is given by:

$$Y_{(\lfloor \beta T \rfloor)} + \frac{\widehat{\sigma}(Y_{(\lfloor \beta T \rfloor)})}{\widehat{\xi}} \left(\left(\frac{1 - \alpha}{1 - \beta} \right)^{-\widehat{\xi}} - 1 \right)$$

where $\widehat{\sigma}(Y_{(\lfloor \beta T \rfloor)})$ and $\widehat{\xi}$ are, for instance, maximum likelihood estimates based on Pickands-Balkema-De Hann's theorem.

Other quantile estimates can be built that avoid the estimation of the scaling factor σ. Among them, the most common is known as the Hill estimator of the Value at Risk:

$$Y_{(\lfloor \beta T \rfloor)} + Y_{(\lfloor \beta T \rfloor)} \left(\left(\frac{1-\alpha}{1-\beta} \right)^{-\widehat{\xi}^{Hill}} - 1 \right)$$

where $\widehat{\xi}^{Hill}$ is the Hill estimator of the tail index (see [46, 48] for more on tail index estimates and for instance the Pickands estimator):

$$\widehat{\xi}^{Hill} = \frac{1}{T - \lfloor \beta T \rfloor + 1} \sum_{j > \lfloor \beta T \rfloor} \log(Y_{(j)}) - \log(Y_{(\lfloor \beta T \rfloor)})$$

These methods rely strongly on the assumption that we are in the tails. In other words the very difficulty in applications of EVT is in the choice of the threshold, or equivalently on the value of β. If the threshold is small, then the approximation based on the Pickands-Balkema-De Hann theorem will be hard to justify and the associated figures irrelevant. On the other hand, if the threshold is too high, then the available data for peaks over threshold will not be sufficiently large to calibrate the parameters. This tradeoff is the main issue in using EVT and theoretical answers have been given in [25].

We have seen how Extreme Value Theory helps to use the entire tail of the distribution in order to estimate the quantile instead of relying on the empirical quantile only.
In the next section we provide another framework that tries to deal directly with the quantiles instead of building an empirical distribution. This approach relies on historical data and uses quantile regressions.

3 Quantile regressions

Quantile regressions have been introduced in the late 70s in a seminal article by Koenker and Basset[58] and consist in regressions in which, contrary to usual OLS regressions that fit the data to the mean, fit the data to a quantile of the observations.
This approach has been extended to estimate Value at Risk, noticeably through the introduction of the so-called CAViaR (Conditional Autoregressive Value at Risk) model by Engle and Manganelli[14] [31]. This paper, along with plenty of other papers that we do not review here (see for instance [22], [79], ...), models the Value at Risk through an autoregressive (potentially

[14]This model appears in the scientific community in 1999.

nonlinear) process, typically with one lag (in [59] they propose a specification with more lagged terms):

$$VaR_{t,t+1} = f_\beta(VaR_{t-1,t}, y_t)$$

where f_β is some explicit function whose parameter β has to be calibrated and where y_t is a new piece of data that brings new information to compute Value at Risk.

Although this approach is more adapted to a portfolio whose constitution does not change with time, we can use it with data from historical approches. In that case, y_t is the P&L we would have had at time t if we had held the portfolio between $t-1$ and t
With this approach, the parameter β is calibrated while minimizing the usual criterion of quantile regression:

$$\min_\beta \sum_{s, y_s \geq -VaR_{s-1,s}} (1-p)|y_s + VaR_{s-1,s}| + \sum_{s, y_s < -VaR_{s-1,s}} p|y_s + VaR_{s-1,s}|$$

This kind of minimization – that can be done with most scientific softwares and uses the simplex method – has to be done after each date to recompute the entire path of Value at Risk. This is the reason why quantile regressions seem not to be well suited to portfolios whose content is varying across time.

As for the use of Extreme Value Theory, this method is semi-parametric. No distributional assumption is indeed made on the data but the way Value at Risk is computed is parametric since it depends on the chosen family of functions $((f_\beta)_\beta)$.
Usual specifications are:

- The baseline CAViaR specification:

$$VaR_{t,t+1} = VaR_{t-1,t} + \beta(1_{y_t \leq -VaR_{t-1,t}} - (1-p))$$

  This specification means that the Value at Risk is updated upward in case y_t is less than the threshold and decreases very slowly otherwise.

- The symmetric absolute value specification:

$$VaR_{t,t+1} = \beta_0 + \beta_1 VaR_{t-1,t} + \beta_2 |y_t|$$

- The asymmetric slope specification:

$$VaR_{t,t+1} = \beta_0 + \beta_1 VaR_{t-1,t} + \beta_2 y_t^+ + \beta_3 y_t^-$$

- The indirect GARCH(1,1) specification:

$$VaR_{t,t+1} = \sqrt{\beta_0 + \beta_1 VaR_{t-1,t}^2 + \beta_2 y_t^2}$$

This approach, in the case of the baseline CAViaR specification, is close to a completely different approach based on a stochastic algorithm that we present in the part on Monte-Carlo simulations.

Part II

Parametric methods for Value at Risk

In the historical approach – as for the Monte Carlo simulations tackled in the next part –, one computes the value of the portfolio for each past value of the risk factors. A natural consequence is that each financial product within the portfolio has to be priced for each new value of the risk factors. Hence computing the Value at Risk with the above methods requires a lot of computations and must, at least theoretically, embed all the pricers in order to continuously reevaluate the portfolio. Although methods will be discussed in the Monte Carlo framework to fasten these time-consuming processes, the pricers must always be used to price the numerous assets within the portfolio for several sets of values for the risk factors. In this part, which is the core part of this document, we present an alternative route to computing an estimation of the Value at Risk of a portfolio that does not embed any asset pricing. More precisely, the parametric methods we are now going to discuss only rely on the current price and Greeks of every asset, all these figures being usually available from current pricing.
In other words, the parametric or analytical methods approximate the value of the portfolio in order to skip the time-consuming reevaluation of the portfolio and leave us with a tradeoff between accuracy and speed. They are much faster than historical or Monte Carlo methods, but not as accurate unless the pricing function can be approximated well by a linear or quadratic function of the risk factors.

The first widespread approach of this kind has been proposed by RiskMetrics^{TM} and we shall present it in the next section. The approximation used being linear, many endeavors have been used to improve it and an important literature has been dedicated to quadratic approximations and the so-called *quadratic portfolios*. This literature with new advances will be discussed below.

4 The original RiskMetrics parametric approach

Originally developed in the 80's by J.P. Morgan for its own risk measurement[15], the RiskMetrics methodology has been widely used since it was made available to all market's participants in 1994 through the publication of a technical document and the free access to an important variance-covariance data set. As a commitment toward transparency in risk measurements, J.P. Morgan published in 1996 a very complete document that is now considered seminal and known as the *RiskMetrics Technical Document* [70]. In what follows we present the original parametric approach as stated in [70] (mainly Parts II and III) and the reader may also refer to the updated document *Return to RiskMetrics: The Evolution of a Standard* [69] published in 2001 that presents the same distributional assumptions concerning risk factors along with other methods to compute Value at Risk (Monte Carlo and historical simulation).

As for all parametric Value at Risk computation methods, the original RiskMetrics approach consists of three main steps:

- First, relevant risk factors have to be chosen with respect to the portfolio and a suitable distributional assumption has to be made in order to describe each risk factor and their interdependence. In the original RiskMetrics framework, risk factors log-returns (or variations for interest rates) are assumed to be conditionally normally distributed (multivariate normal distribution), the conditionality being on the variance-covariance matrix of the returns. We will detail below this hypothesis but it is important to notice that it naturally incorporates two main stylized facts of financial time series: volatility clustering and fat-tailed distributions of returns.

  The risk factors are usually stock prices, stock indices, interest rates for different time horizons (discrete yield curve), foreign exchange rates and commodity prices[16].

- Second, the portfolio under scrutiny must be represented in terms of the risk factors. In other words the exposition to the risk factors must be computed. Mathematically, this step consists in writing the expression of the loss function $\mathcal{L}_{t,t+1}(X^1, \ldots, X^n)$ and/or of the return function $\mathcal{R}_{t,t+1}(X^1, \ldots, X^n)$. In practice, this exercise depends on the nature

[15] Spun off from J.P. Morgan in 1998, the RiskMetrics Group was acquired in 2010 by MSCI.
[16] In more general settings, volatility indicators or credit spreads can be incorporated.

of the portfolio. For an equity-only portfolio, the function $\mathcal{R}_{t,t+1}(X^1, \ldots, X^n)$ will be a linear function of stocks' returns $(r^1_{t,t+1}, \ldots, r^n_{t,t+1})$[17] (and exchange rates, if needed). The situation is more complex for an equity derivative portfolio since, the payoffs of the assets being nonlinear, the function $\mathcal{R}_{t,t+1}$ has to take account of the convexity of the different assets it includes. In practice, for parametric computations of the Value at Risk, the resulting function will be approximated by a linear function of the risk factors' returns (Delta-normal approach) or as documented in [70] by a quadratic function of these returns (Delta-Gamma(-Theta) approach[18]).

Now, for a fixed income portfolio, the first step is to map the cash flows in terms of the selected time horizons and the same problem as above appears with nonlinear derivatives (for more details on the fixed-income specific approach, see Part III of [70] and Chapter 5 of [69]).

- Third, the computation of the Value at Risk is done using statistical results on gaussian variables when the function $\mathcal{R}_{t,t+1}$ is linear in the risk factors' returns (Delta-normal approach) or using one of the numerous methods presented below in the nonlinear case when a Delta-Gamma-(Theta) approach is used (see below)

Here we see the main issues that one faces when building a parametric method to compute Value at Risk: the combination of the distributional assumptions on the risk factors and the approximation of the portfolio valuation must lead to an approximation of the portfolio return distribution that is tractable for quantiles estimation.

Let us start now with the description of the distributional assumptions on the risk factors' returns made in the original RiskMetrics framework.

4.1 Distributional assumptions underlying the RiskMetrics approach

The RiskMetrics model for the distribution of the evolution of the risk factors is based on the assumption that log-returns of prices (or variations in the case of interest rates) are independent

[17] Log-returns being used in RiskMetrics, a linear approximation is used to end up with a linear function.

[18] This approach will be presented in the next subsection and we will review the different methods to estimate quantiles of the so-called quadratic portfolios

across time and normally distributed, when appropriately scaled by an appropriate measure of volatility.

If we denote $(X_t^1, \ldots, X_t^n)_t$ the process of the risk factors (here we assume that these risk factors are prices but the approach would be *mutatis mutandis* the same with interest rates) the log-returns between time t and time $t+1$ are assumed to be

$$r_{t,t+1}^i = \sigma_t^i \epsilon_{t,t+1}^i$$

where, the distribution of $\epsilon_{t,t+1} = (\epsilon_{t,t+1}^1, \ldots, \epsilon_{t,t+1}^n)$ is gaussian with $\langle \epsilon_{t,t+1}^i, \epsilon_{t,t+1}^j \rangle = \rho_t^{ij}$ conditionally on the realization of the variance covariance matrix process $(\Sigma_t)_t$ where $\Sigma_t^{ij} = \sigma_t^i \sigma_t^j \rho_t^{ij}$.

More precisely, we assume a multivariate IGARCH(1,1) model for the returns, namely:

$$\begin{aligned} r_{t,t+1}^i &= \sigma_t^i \epsilon_{t,t+1}^i \\ {\sigma_t^i}^2 &= \lambda {\sigma_{t-1}^i}^2 + (1-\lambda) {r_{t-1}^i}^2 \\ \rho_t^{ij} &= \frac{1}{\sigma_t^i \sigma_t^j} \left[\lambda \sigma_{t-1}^i \sigma_{t-1}^j \rho_{t-1}^{ij} + (1-\lambda) r_{t-1}^i r_{t-1}^j \right] \end{aligned}$$

where $\epsilon_{t,t+1} | \Sigma_t$ is distributed as $\mathcal{N}(0, \Sigma_t)$ and where the variables $\epsilon_{t,t+1} | \Sigma_t$ are independent across time.

Noticeably, no drift is assumed for the risk factors and hence the returns have a zero mean.

This IGARCH(1,1) assumption on the returns is based on common stylized facts: no drift for short-to-medium time horizons (less than 3 months), no autocorrelations between returns, positive autocorrelations between squared returns (volatility clustering), fat tails, ...

Another way to think about the volatility and correlation processes is to see them as exponentially weighted moving average (EWMA) estimates. Then, λ scales the relevant number of observations in the past to be taken into account. In practice, RiskMetrics considers $\lambda = 0.94$ for one-day returns and $\lambda = 0.97$ for one-month returns.

4.2 A first approximation of the portfolio: the Delta-Normal approach

The general idea underlying Value at Risk parametric computation methods is to approximate the return of the portfolio using approximations of the pricing formulas of each asset in the portfolio in order to end up with a simple analytic formula for the Value at Risk. In the following paragraphs, we present the most simple approximation which is the linear approximation. In a word, this approximation is exact for linear instruments[19] and is however arguable when the portfolio is more general. As an example, a portfolio containing options will be considered a portfolio a stocks where each option is replaced by a position consisting of Δ shares of the underlying stock (the name Delta-Normal coming indeed from the Greeks in option pricing). Mathematically, if we denote V the value of the portfolio, then[20]:

$$V(X_{t+1}^1, \ldots, X_{t+1}^n) \simeq V(X_t^1, \ldots, X_t^n) + \sum_{i=1}^{n} \frac{\partial V}{\partial X^i}(X_t^1, \ldots, X_t^n)(X_{t+1}^i - X_t^i)$$

$$V(X_{t+1}^1, \ldots, X_{t+1}^n) \simeq V(X_t^1, \ldots, X_t^n) + \sum_{i=1}^{n} \frac{\partial V}{\partial X^i}(X_t^1, \ldots, X_t^n) r_{t,t+1}^i X_t^i$$

Hence,

$$R = \frac{V(X_{t+1}^1, \ldots, X_{t+1}^n) - V(X_t^1, \ldots, X_t^n)}{V(X_t^1, \ldots, X_t^n)}$$

is approximated by:

$$\tilde{R} = \sum_{i=1}^{n} \frac{X_t^i}{V_t} \frac{\partial V}{\partial X^i}(X_t^1, \ldots, X_t^n) r_{t,t+1}^i$$

This function is indeed linear in the returns of the risk factors and we can find the Value at Risk of the associated *linear portfolio*. This is the main strength of the RiskMetrics Delta-Normal method. Since the returns of the risk factors are assumed to be conditionally normal, once the exponentially weighted moving average has been calculated, $\tilde{R}$ "becomes" a random variable with normal distribution, the mean being obviously 0 and the standard deviation being:

$$\sigma_t^{\tilde{R}} = \sqrt{\sum_{i,j=1}^{n} \left(\frac{X_t^i}{V_t} \frac{\partial V}{\partial X^i}(X_t^1, \ldots, X_t^n) \right) \Sigma_t^{ij} \left(\frac{X_t^j}{V_t} \frac{\partial V}{\partial X^j}(X_t^1, \ldots, X_t^n) \right)}$$

[19] An approximation is done when returns are supposed to be log-returns.

[20] assuming returns instead of log-returns, although this may not be practical top deal with different time horizons

Now, using this approximated distribution for the portfolio returns, the estimation of the Value at Risk estimate boils down to the computation of the $(1-p)$-quantile of a random variable with gaussian distribution. This can be calculated easily using the quantile z_p of a standard gaussian variable and we obtain the expression (where the Value at Risk is, by convention, positive for losses and expressed in terms of portfolio return and not as a value[21]):

$$\widehat{VaR}(p)_{t,t+1} = -z_{1-p}\sqrt{\sum_{i,j=1}^{n}\left(\frac{X_t^i}{V_t^i}\frac{\partial V}{\partial X^i}(X_t^1,\ldots,X_t^n)\right)\Sigma_t^{ij}\left(\frac{X_t^j}{V_t}\frac{\partial V}{\partial X^j}(X_t^1,\ldots,X_t^n)\right)}$$

$$\widehat{VaR}(p)_{t,t+1} = z_{p}\sqrt{\sum_{i,j=1}^{n}\left(\frac{X_t^i}{V_t^i}\frac{\partial V}{\partial X^i}(X_t^1,\ldots,X_t^n)\right)\Sigma_t^{ij}\left(\frac{X_t^j}{V_t}\frac{\partial V}{\partial X^j}(X_t^1,\ldots,X_t^n)\right)}$$

Using the analogy with the usual notations in option pricing, this equation is often written as:

$$\boxed{\widehat{VaR}(p)_{t,t+1} = z_p\sqrt{\textstyle\sum_{i,j=1}^{n}\left(\frac{X_t^i}{V_t^i}\Delta^i\right)\Sigma_t^{ij}\left(\frac{X_t^j}{V_t}\Delta^j\right)}, \qquad \Delta = \nabla V(X_t^1,\ldots,X_t^n)}$$

N.B. : We provide below a table of the useful values of the quantiles

p	z_p
95%	1.645
99%	2.326
99.9%	3.090

For instance, if we consider a portfolio of stocks with proportions $\alpha^1,\ldots,\alpha^n$ of the wealth invested respectively in stocks $1,\ldots,n$, then, $\frac{X_t^i}{V_t}\frac{\partial V}{\partial X^i}(X_t^1,\ldots,X_t^n)$ is simply the weight α^i and the Value at Risk estimate $\widehat{VaR}(p)_{t,t+1}$ is:

$$\widehat{VaR}(p)_{t,t+1} = z_p\sqrt{\sum_{i,j=1}^{n}\alpha^i\alpha^j\Sigma_t^{ij}}$$

In this stock-only example, the Delta-Normal approach works well as long as the initial distributional assumptions on the risk factors fit the data. However, if we consider a portfolio containing options, the approximation is less accurate and can even be completely wrong if one considers a Δ-hedged portfolio of options. In that case, the Delta-Normal approach provides a

[21] This will be the case throughout the paper.

very inaccurate answer since the portfolio appears to be risk-free and the Value at Risk estimation is not even well defined and will be estimated to be equal to nought.

Subsequently, other parametric methods have been developed to take account of nonlinear instruments (Stock options, Bond options, FX options) and these methods are often referred to as Delta-Gamma-Normal or Delta-Gamma-Theta-Normal approaches. These approaches were not part of the initial RiskMetrics methodology but Zangari [85], from RiskMetrics, proposed one of them and we present in the next paragraphs the basic ideas of the Delta-Gamma-(Theta)-Normal approach.

4.3 Introduction to the Delta-Gamma-(Theta)-Normal approach

To go beyond the linear approximation introduced in the above paragraphs, and find more accurate approximations for the Value at Risk, we need to go further in the Taylor expansion started above for the value of the portfolio. If we consider the first two terms arising from the change in the risk factors, keeping the terms appearing in Itô's formula and for which the traders usually have figures straightforwardly available, we get:

$$V(X_{t+1}^1, \ldots, X_{t+1}^n) \simeq V(X_t^1, \ldots, X_t^n) + \Theta \times 1 + \sum_{i=1}^{n} \Delta^i r_{t,t+1}^i X_t^i + \frac{1}{2} \sum_{i,j=1}^{n} \Gamma^{ij} X_t^i X_t^j r_{t,t+1}^i r_{t,t+1}^j$$

where

$$\Theta = \frac{\partial V}{\partial t} \qquad \Delta^i = \frac{\partial V}{\partial X^i} \qquad \Gamma^{ij} = \frac{\partial^2 V}{\partial X^i \partial X^j}$$

The Theta term ($\Theta \times 1$) takes into account the evolution of the portfolio value with time (as for the time value of an option) and is often neglected in Value at Risk computations since the time horizon is often small.

Ignoring from now the Theta term and focus therefore on the Delta-Gamma approach, we get for the return of the portfolio:

$$R \simeq \tilde{R} = \sum_{i=1}^{n} \frac{X_t^i}{V_t} \Delta^i r_{t,t+1}^i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{X_t^i}{V_t} \frac{X_t^j}{V_t} \left(V_t \Gamma^{ij} \right) r_{t,t+1}^i r_{t,t+1}^j$$

Hence, to compute the Value at Risk using this approximation, we have to compute the quantile of a distribution which is no longer gaussian. In particular, in spite of the zero-mean return assumption on the risk factors, the portfolio return approximation $\tilde{R}$ exhibits a non-zero mean corresponding to the Γ terms, a different variance, a distribution that may be skewed and a different kurtosis that quantifies to what extent the distribution is heavy-tailed.

In the next section, we will review the different approaches proposed in the literature to approximate the quantiles of a random variable which is a quadratic polynomial of gaussian variables. An important literature has indeed been dedicated to this problem which is, in our financial context, often referred to as the computation of Value at Risk for quadratic portfolios. We will also see later in this text that this approach may allow for more general distributional assumptions for the factors, but from now we stick to the Delta-Gamma-Normal approach[22].

5 The Delta-Gamma-Normal approach

The Delta-Gamma approach was introduced in the preceding subsection to approximate the return of a portfolio involving nonlinear instruments. This approximation consists in fact, with the conditional normality assumptions of RiskMetrics – and in that case we call it Delta-Gamma-Normal –, in approximating the distribution of the portfolio return by the distribution of a quadratic multivariate polynomial in gaussian variables:

$$\tilde{R} = \sum_{i=1}^{n} \frac{X_t^i}{V_t} \Delta^i r_{t,t+1}^i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{X_t^i}{V_t} \frac{X_t^j}{V_t} \left(V_t \Gamma^{ij} \right) r_{t,t+1}^i r_{t,t+1}^j$$

Hence the problem boils down to the computation of such a distribution's quantiles and several methods have been developed for that purpose:

- The use of a **gaussian approximation** for the distribution of $\tilde{R}$ and the subsequent determination of the quantile on the fitted normal distribution. This method ends up to an approximation of the Value at Risk involving the first two moments of $\tilde{R}$.

[22] Taking into account the time value of the portfolio, i.e. Θ, and hence extending to the Delta-Gamma-Theta-Normal approach will never be an issue.

- The use of **Johnson transformation**. It consists in choosing a distribution among the Johnson family of distributions in order to fit the first four moments of $\tilde{R}$. The advantage of the Johnson distributions is that quantiles of these distributions only depends on the quantiles of a standard normal random variable.

- The use of **Cornish-Fisher expansion**. As in the case of the Johnson transformations, the idea is to use the moments of $\tilde{R}$ to derive an approximation of the Value at Risk. The Cornish-Fisher expansion is indeed an approximation of the quantiles of a distribution using polynomials in the quantiles of a gaussian distribution with coefficients depending on the moments of the distribution under scrutiny.

- The use of **Solomon-Stephens approximation**. $\tilde{R}$ can be shown to be a linear combination of independent non-central χ^2 random variables and some authors (see for instance [14]), following Solomon-Stephens paper [78], proposed to approximate this variable with a variable $K_1 X^{K_2}$ where K_1 and K_2 are two constants and where X follows a χ^2 distribution with a degree of freedom to be calibrated[23].

- The use of **saddle point approximations**. Using the fact that $\tilde{R}$ can be represented as the sum of polynomials of degree two in independent random normal variables, this method starts with the determination of an analytical expression for the moment generating function. Then, using the cumulant generating function, a saddle point approximation is used to approximate the tail function and deduce the Value at Risk (see [34] as an instance).

- The use of **Fourier transform**. The characteristic function of $\tilde{R}$ can be found analytically using the representation of $\tilde{R}$ as a sum of independent non-central χ^2 random variables. Hence, using Fast Fourier Transform, one can go back to an estimate of the probability distribution function of $\tilde{R}$ and thus the tail function and the Value at Risk

[23] It is not clear that this method can indeed be used in practice. The methodology described to calibrate K_1, K_2 and the degree of freedom is indeed arguable.

(see [1, 76]).

- The use of an approximation based on the concept of **principal component Value at Risk** introduced in [16][24]. This approach, which has not be designed for a portfolio with a positive Γ – but this case is clearly not of great interest when dealing with Value at Risk, at least in the short run, as long as the quadratic approximation is valid –, consists of a closed-form approximation of the cumulative distribution function. This approximation, along with a lower bound and an upper bound in some cases, is based on an asymptotic approximation of integrals over quadrics.

- The use of **partial Monte-Carlo simulation** to estimate the quantile. This method is not a parametric method but it is not a pure Monte-Carlo approach either since the value of the portfolio is not recomputed after each draw. Rather, the (partial) Monte-Carlo simulation is only used to compute the quantile of the parametric Delta-Gamma approximation of the return, namely $\tilde{R}$.

5.1 Gaussian approximation

The Delta-Gamma-Normal approach provides an approximation of the portfolio return distribution and we can write analytically the first two moments of this distribution. Namely, if we denote $\tilde{\Delta}^i = \frac{X_t^i}{V_t}\Delta^i$ and $\tilde{\Gamma}^{ij} = \frac{X_t^i}{V_t}\frac{X_t^j}{V_t}(V_t\Gamma^{ij})$ ($\tilde{\Delta}$ and $\tilde{\Gamma}$ being respectively the associated vector and matrix) then we can straightforwardly show that the mean and the variance of the quadratic portfolio's return are:

$$\mu = \mathbb{E}[\tilde{R}] = \frac{1}{2}Tr(\tilde{\Gamma}\Sigma)$$

and

$$\sigma^2 = \mathbb{E}\left[\left(\tilde{R} - \mathbb{E}[\tilde{R}]\right)^2\right] = \tilde{\Delta}'\Sigma\tilde{\Delta} + \frac{1}{2}Tr\left((\tilde{\Gamma}\Sigma)^2\right)$$

Hence, the basic idea of the gaussian approximation is to fit a normal distribution to (μ, σ) and to deduce an estimate for the quantile at stake.

[24] Most of the results in this paper seem to suffer from typos but the approach is relevant. In this paper, we present corrected results in a similar fashion as in [17].

In other words, we estimate the Value at Risk by:

$$\widehat{VaR}_{t,t+1}(p) = -\mu - \sigma z_{1-p}$$

$$\widehat{VaR}_{t,t+1}(p) = -\mu + \sigma z_p$$

This approximation is obviously crude and better ones have been proposed, using for instance more moments. This is the the case of the Johnson transformation approach, the Cornish-Fisher approximation and the Solomon Stephens approach that are presented below.

5.2 Johnson transformation

The Johnson transformation approach proposed in [70] consists in finding a distribution that has the same first four moments $\tilde{R}$ and whose quantiles can easily be calculated.
To understand this approach, let us recall that the first four moments of $\tilde{R}$ are different from those of a normal random variable. Hence, we have to consider a large family of distributions to be able to fit the first four moments. In addition, any distribution of this family must have quantiles that are easy to compute. In fact, the approach employed consists in fitting a distribution amongst the Johnson family of distributions (see [47, 68, 70]) on the first four moments of $\tilde{R}$. The advantage of these Johnson distributions is that they are obtained through the distribution of monotonic transformations of normal random variables and hence the quantiles are easily computed as analytical functions of the gaussian quantiles $(z_p)_p$.

More precisely $\tilde{R}$ distribution is approximated by the distribution of $Y = \mathcal{T}(Z)$ where Z is a standard normal random variable and where g is of the form:

$$\mathcal{T}(z) = a + bg\left(\frac{z-c}{d}\right)$$

for g a certain monotonic function where (a, b, c, d) and the function g are calibrated according to the algorithm developed in [47].

Once this calibration has been done, the Value at Risk is estimated by:

$$\widehat{VaR}_{t,t+1}(p) = -\mathcal{T}(z_{1-p})$$

This approximation was proposed quite early in the mid-90s but does not provide accurate results for Value at Risk. Other approaches using moments have also been proposed and we now turn to the most famous of them: the use of Cornish-Fisher approximation.

5.3 Cornish-Fisher approximation

To compute the quantile of a multivariate polynomial in gaussian variables, one can use what is usually called the Cornish-Fisher expansion. This approximation consists in evaluating the quantiles of a random variable using its moments and the corresponding quantiles of a standard gaussian variable.

This approach is different from the Johnson transformation approach since no parametric assumption is used to compute the quantile once the moments of $\tilde{R}$ have been taken into account. This formula is often used with the first four moments.

Let us denote indeed

$$\mu = \mathbb{E}[\tilde{R}]$$

$$\sigma^2 = \mathbb{E}\left[\left(\tilde{R} - \mathbb{E}[\tilde{R}]\right)^2\right]$$

$$s = \frac{\mathbb{E}\left[\left(\tilde{R} - \mathbb{E}[\tilde{R}]\right)^3\right]}{\sigma^3}$$

and

$$\kappa = \frac{\mathbb{E}\left[\left(\tilde{R} - \mathbb{E}[\tilde{R}]\right)^4\right]}{\sigma^4} - 3$$

respectively the mean, the variance, the skewness and the kurtosis of $\tilde{R}$.

These four parameters can be calculated analytically using standard manipulations on gaussian variables[25] and the Cornish-Fisher approximation is:

[25] Using the same notations as above we obtain indeed the following expressions for the skewness and kurtosis:

$$s = \frac{1}{\sigma^3}\left[3\tilde{\Delta}'\Sigma\tilde{\Gamma}\Sigma\tilde{\Delta} + Tr\left((\tilde{\Gamma}\Sigma)^3\right)\right]$$

$$\kappa = \frac{1}{\sigma^4}\left[12\tilde{\Delta}'\Sigma\tilde{\Gamma}\Sigma\tilde{\Gamma}\Sigma\tilde{\Delta} + 3Tr\left((\tilde{\Gamma}\Sigma)^4\right)\right]$$

$$\widehat{VaR}_{t,t+1}(p) = -\left(\mu + \sigma\left(z_{1-p} + (z_{1-p}^2 - 1)\frac{s}{6} + (z_{1-p}^3 - 3z_{1-p})\frac{\kappa}{24} - (2z_{1-p}^3 - 5z_{1-p})\frac{s^2}{36}\right)\right)$$

$$\widehat{VaR}_{t,t+1}(p) = -\mu + \sigma\left(z_p - (z_p^2 - 1)\frac{s}{6} + (z_p^3 - 3z_p)\frac{\kappa}{24} - (2z_p^3 - 5z_p)\frac{s^2}{36}\right)$$

This expression can in fact be seen as an inversion of the Edgeworth expansion and can therefore be computed for any number of moments instead of the usual four moments as above[26]. In particular, this methodology is equivalent, when using only two moments, to fitting a normal distribution to the distribution of $\tilde{R}$.

Now, this technique has been used and compared using different numbers of moments [53], [68], [71] and the main advantage of this methodology in addition to its robustness compared to the Johnson transformation technique is that it is very fast. Using the Cornish-Fisher approximation indeed only requires the computation of the first moments of $\tilde{R}$. Hence, after the Δ and Γ of the portfolio have been computed, the problem is limited to the computation of moments for which analytical expressions are available and consists, in terms of computational complexity, of sums and products of matrices whose sizes are bounded by the number of factors.

Now, the Cornish-Fisher approximation has drawbacks. First, although this seems not to happen for practical values of the skewness and kurtosis coefficients, the above polynomial is not necessarily monotone. Thus, we may theoretically end up with a situation in which the estimation of VaR_p with $p = 95\%$ happens to be greater than the estimation of VaR_p with $p = 99\%$. Although this example is unlikely to occur in practice, the very simple fact that the approximation is not monotone is an issue.

Moreover, the precision of Cornish-Fisher approximation is not increasing with the number of moments used to compute the approximation[27].
Finally, another problem arises when the Value at Risk is computed for very large values of

[26]For a proof of the formula and the techniques used to compute the coefficients, the interested reader may look at [53].

[27]This is based on the fact that the Edgeworth expansion underlying Cornish-Fisher approximation may not converge where needed for all the distributions at stake.

p. In that case indeed, since the Cornish-Fisher approximation consists in approximating the true quantile of $\tilde{R}$ using a polynomial of degree d in the quantile of a standard normal variable, then the Cornish-Fisher approximation consists in fact in approximating the random variable $\tilde{R}$ by a random variable of the form $P(N)$ where P is a polynomial of degree d that we assume to be increasing (see the first issue described above when this assumption is not verified) and N a standard normal random variable. Hence, the probability distribution function of $\tilde{R}$ is approximated by a probability distribution function of the form:

$$\frac{1}{P'(P^{-1}(x))} \frac{\exp\left(-\frac{P^{-1}(x)^2}{2}\right)}{\sqrt{2\pi}}$$

The consequence is that the Cornish-Fisher approximation may not be accurate in the tails since the tail approximation depends strongly on the degree of P. In practice, this approximation should not be used for very high values of p ($p = 99.9\%$ for instance).

5.4 Solomon-Stephens approximation

Another method that uses the moments of the distribution of $\tilde{R}$ is the Solomon-Stephens approximation. This method, along with the methods based on saddle point approximations or Fourier transforms (see below), is based on a decomposition of $\tilde{R}$ as the sum of independent random variables following non-central χ^2 distributions. This decomposition is going to be presented hereafter and it should be noticed that in this method we do not need to compute the actual decomposition.

The Delta-Gamma-Normal approximation of the portfolio return is:

$$\tilde{R} = \sum_{i=1}^{n} \frac{X_t^i}{V_t} \Delta^i r_{t,t+1}^i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{X_t^i}{V_t} \frac{X_t^j}{V_t} \left(V_t \Gamma^{ij}\right) r_{t,t+1}^i r_{t,t+1}^j$$

or in a more compact fashion:

$$\tilde{R} = r'_{t,t+1} \tilde{\Delta} + \frac{1}{2} r'_{t,t+1} \tilde{\Gamma} r_{t,t+1}$$

where $r_{t,t+1} \sim \mathcal{N}(0, \Sigma)$.

We consider first a Cholesky decomposition of Σ:

$$\Sigma = T_\Sigma T_\Sigma'$$

Then, we consider a diagonalization of $T_\Sigma' \tilde{\Gamma} T_\Sigma$ in an orthonormal basis:

$$T_\Sigma' \tilde{\Gamma} T_\Sigma = \Omega D \Omega'$$

Introducing $Z = \Omega' T_\Sigma^{-1} r_{t,t+1}$, we have that:

$$\tilde{R} = Z' \Omega' T_\Sigma' \tilde{\Delta} + \frac{1}{2} Z' D Z$$

where $Z \sim \mathcal{N}(0, I)$.
Hence, $\tilde{R}$ can be written as

$$\tilde{R} = \sum_j c_j Z_j + \frac{1}{2} d_j Z_j^2$$

where $(Z_j)_j$ is a family of independent gaussian random variables and where the families $(c_j)_j$ and $(d_j)_j$ are families of constants depending only on $\tilde{\Delta}$, $\tilde{\Gamma}$ and Σ.

Another way to write this decomposition is:

$$\tilde{R} = \sum_j \frac{1}{2} d_j \left(Z_j + \frac{c_j}{d_j} \right)^2 - \sum_j \frac{c_j^2}{2 d_j}$$

Hence, $\tilde{R} + \sum_j \frac{c_j^2}{2 d_j}$ is the sum of independent random variables following a non-central χ^2 distribution.

This last representation is used in Solomon-Stephens approximation. Solomon and Stephens [78] indeed suggested that the sum of independent non-central χ^2 variables can be approximated by a random variable $K_1 X^{K_2}$ where K_1 and K_2 are two constants and where X follows a χ^2 distribution with ν degrees of freedom, where K_1, K_2 and ν are calibrated using the method of moments, *i.e.* to match the first three moments of $\tilde{R} + \sum_j \frac{c_j^2}{2 d_j}$.
On the one hand, the first three moments of $\tilde{R} + \sum_j \frac{c_j^2}{2 d_j}$ can be computed using the first three moments of $\tilde{R}$:

$$\mathbb{E}[\tilde{R}] = \frac{1}{2}Tr(\tilde{\Gamma}\Sigma)$$

$$\mathbb{E}[\tilde{R}^2] = \left[\frac{1}{2}Tr(\tilde{\Gamma}\Sigma)\right]^2 + \tilde{\Delta}'\Sigma\tilde{\Delta} + \frac{1}{2}Tr\left((\tilde{\Gamma}\Sigma)^2\right)$$

$$\mathbb{E}[\tilde{R}^3] = \left[\frac{1}{2}Tr(\tilde{\Gamma}\Sigma)\right]^3 + \frac{3}{2}\left[\tilde{\Delta}'\Sigma\tilde{\Delta} + \frac{1}{2}Tr\left((\tilde{\Gamma}\Sigma)^2\right)\right]Tr(\tilde{\Gamma}\Sigma) + 3\tilde{\Delta}'\Sigma\tilde{\Gamma}\Sigma\tilde{\Delta} + Tr\left((\tilde{\Gamma}\Sigma)^3\right)$$

On the other hand, we can write the first three moments of $K_1X^{K_2}$ which are respectively:

$$(K_1)^m 2^{mK_2}\frac{\Gamma(sK_2 + \frac{\nu}{2})}{\Gamma(\frac{\nu}{2})}$$

for m being 1, 2 or 3.

The problem is then to find K_1, K_2 and r and we are left with a nonlinear system of three equations with three variables.

Techniques to solve this system are explained in [78, 14][28]. Once K_1, K_2 and ν have been calibrated, the Value at Risk estimate is expressed in terms of the $(1-p)$-quantile of a χ^2_ν random variable $q^{\chi^2_\nu}_{1-p}$ which is tabulated:

$$\widehat{VaR}_{t,t+1}(p) = -\left(K_1\left(q^{\chi^2_\nu}_{1-p}\right)^{K_2} - \sum_j \frac{c_j^2}{2d_j}\right)$$

$$\widehat{VaR}_{t,t+1}(p) = -K_1\left(q^{\chi^2_\nu}_{1-p}\right)^{K_2} + \sum_j \frac{c_j^2}{2d_j}$$

5.5 Saddle point approximation

As we have seen above, $\tilde{R}$ is the sum of polynomials of degree two in independent random normal variables:

$$\tilde{R} = \sum_j c_j Z_j + \frac{1}{2}d_j Z_j^2$$

where the Z_j's are independent standard normal variables and where the c_j's and d_j's are constants that can be computed using linear algebra on Σ and $\tilde{\Gamma}$.

[28] It is not clear that a perfect fit can be obtained in the sense that none of the papers prove that the system is indeed invertible.

The reader must indeed notice that, in practice, this decomposition only requires linear algebra, namely a Cholesky decomposition and a diagonalization. Hence, it can be carried out numerically even for a large number of risk factors.

From this, we can calculate the moment generating function of $\tilde{R}$ analytically:

$$M(t) = \mathbb{E}\left[e^{t\tilde{R}}\right] = \prod_j \mathbb{E}\left[e^{t(c_j Z_j + \frac{1}{2} d_j Z_j^2)}\right]$$

Hence, for $|t| < \frac{1}{\max_j d_j}$:

$$M(t) = \prod_j \frac{1}{\sqrt{1 - t d_j}} \exp\left(\frac{1}{2} \frac{t^2 c_j^2}{1 - t d_j}\right)$$

$$M(t) = \frac{1}{\sqrt{\prod_j (1 - t d_j)}} \exp\left(\frac{1}{2} \sum_j \frac{t^2 c_j^2}{1 - t d_j}\right)$$

In terms of the initial vector $\tilde{\Delta}$ and matrices $\tilde{\Gamma}$ and Σ, this expression is simply:

$$M(t) = \frac{1}{\sqrt{\det(I - t\Sigma\tilde{\Gamma})}} \exp\left(\frac{1}{2} t^2 \tilde{\Delta}'(\Sigma^{-1} - t\tilde{\Gamma})^{-1}\tilde{\Delta}\right)$$

Now, the cumulant generating function $K(t) = \log(M(t))$ is simply given by

$$K(t) = -\frac{1}{2}\log(\det(I - t\Sigma\tilde{\Gamma})) + \frac{1}{2} t^2 \tilde{\Delta}'(\Sigma^{-1} - t\tilde{\Gamma})^{-1}\tilde{\Delta}$$

The saddle point approximation then consists in approaching the cumulative distribution function of $\tilde{R}$ using the cumulant generating function, by:

$$F_{\tilde{R}}(x) = \Phi(v) + \phi(v)\left(\frac{1}{v} - \frac{1}{u}\right)$$

where

$$v = \frac{y}{|y|}\sqrt{2}\sqrt{xy - K(y)}, \qquad u = y\sqrt{K''(y)}$$

with $K'(y) = x$.

This approximation is rooted to a paper by Lugannani and Rice [63] that, contrary to here, approximates the cumulative distribution function of the sum of i.i.d variables.

5.6 Fourier Transform

Similar to the saddle point approximation, we can opt for an approximation of the cumulative distribution function of $\tilde{R}$ using the classical tools of Fourier transforms. Using the decomposition of $\tilde{R}$ as the sum of polynomials of degree two in independent random variables, we can write analytically the expression of the characteristic function:

$$\psi(\xi) = \mathbb{E}[e^{i\xi\tilde{R}}] = M(i\xi) = \prod_j \frac{1}{\sqrt{1 - i\xi d_j}} \exp\left(-\frac{1}{2}\frac{\xi^2 c_j^2}{1 - itd_j}\right)$$

where the square roots are taken so as to have positive real part (here we assumed that the moment generating function existed in a neighborhood of 0).
Now, using Fourier inversion, we know that the probability distribution function of $\tilde{R}$ is:

$$f_{\tilde{R}}(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\xi x}\psi(\xi)d\xi$$

This integral can be calculated using Fast Fourier Transform and the Value at Risk can be estimated once the cumulated distribution function is reconstructed using numerical integration.

Another slightly different but more direct approach consists in directly writing the cumulative distribution function of the quadratic portfolio's return as:

$$F_{\tilde{R}}(x) = \frac{1}{2} - \frac{1}{\pi}\int_0^{+\infty} \Im\left[f(\xi)e^{-i\xi x}\right]\frac{d\xi}{\xi}$$

for a sufficiently small positive η. Then, the Value at Risk is easily estimated by:

$$1 - p = \frac{1}{2} - \frac{1}{\pi}\int_0^{+\infty} \Im\left[f(\xi)e^{i\xi\widehat{VaR}_{t,t+1}(p)}\right]\frac{d\xi}{\xi}$$

As the method based on saddle point approximation, the Fourier approach uses the fact that we have a closed-form formula for the moment generating function/characteristic function. However, the two methods are very different.
The saddle point approximation indeed only uses functions of a real variable while the Fourier approach deals with complex numbers. The advantage of the Fourier approach is that existing numerical tools can be used while the saddle point approximation is not implemented in most libraries.

More generally, to compute the Value at Risk, if we do not consider the special case of the Delta-Gamma-Normal approximation, the saddle point approximation requires the existence of the moment generating function (and for instance a multivariate Student t-distribution does not have one) while the Fourier transform can be applied to more general random variables. However, the saddle point approximation has an advantage over the use of Fourier Transform if the moment generating functions is not known in closed-form. The moment generating function can indeed be approximated using a few moments and the method applied on the approximated moment generating function may provide acceptable figures (see [34]). The same is not true for the characteristic function, in the case of the Fourier approach, especially when it comes to estimating the probability distribution function or the cumulative distribution function in the tails.

5.7 Principal component Value at Risk

Another approach has been developed in the recent years, starting in 2002 with a paper entitled "Principal component Value at Risk" [16]. In this approach, we consider the same spectral decomposition as the one used to obtain the decomposition of $\tilde{R}$ as a sum of polynomials of degree two in independent normal random variables.
More precisely, if we write $\tilde{R}$ in a *canonical form* as above

$$\tilde{R} = \sum_j c_j Z_j + \frac{1}{2} d_j Z_j^2 = Z'c + \frac{1}{2} Z'DZ$$

with $Z \sim \mathcal{N}(0, I)$, then the Value at Risk linked to the Delta-Gamma-Normal approximation is implicitly given by:

$$\int_{z'c+\frac{1}{2}z'Dz \leq -VaR_p} \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{|z|^2}{2}\right) dz = 1 - p$$

Hence,

$$\int_{\frac{1}{2}(z+D^{-1}c)'D(z+D^{-1}c)-\frac{1}{2}c'D^{-1}c \leq -VaR_p} \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}|z|^2\right) dz = 1 - p$$

$$\int_{\frac{1}{2}(z+D^{-1}c)'D(z+D^{-1}c) \leq -VaR_p+\frac{1}{2}c'D^{-1}c} \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}|z|^2\right) dz = 1 - p$$

$$\int_{\frac{1}{2}z'Dz\leq -VaR_p+\frac{1}{2}c'D^{-1}c} \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\left|z-D^{-1}c\right|^2\right) dz = 1-p$$

Now, we are assuming that we can reorder the terms in the diagonal matrix D so that the smallest eigenvalue is negative and of multiplicity 1:

$$D = diag\left(-d_1^-, -d_2^-, \ldots, -d_{n^-}^-, d_1^+, \ldots, d_{n^+}^+\right)$$

with:

$$-d_1^- < -d_2^- \leq \ldots \leq -d_{n^-}^- \leq 0 \leq d_1^+ \leq \ldots \leq d_{n^+}^+, \qquad n^- + n^+ = n$$

These hypotheses are important to comment. First, assuming that at least one eigenvalue is negative means that we exclude the case of a portfolio with a positive Γ. In practice, such a quadratic portfolio with a positive Γ will never have a Value at Risk greater than the Value at Risk estimated using the Delta-Normal approach. The estimation difficulties arise indeed when the Gamma terms induce additional risk of losses.

Now, concerning the hypothesis on the smallest eigenvalue being simple, this hypothesis is important and means empirically that the asymptotic estimate we will end up with is less accurate when there is a clustering of eigenvalues near the smallest one (in this case of eigenvalues clustering [16] developed an estimate that will not be presented in this text[29]).

Under the assumptions made above we can write any vector y as $y^+ + y^-$ where y^- is the projection on the space spanned by the eigenvectors associated with negative eigenvalues and where y^+ is the projection on the space spanned by the eigenvectors associated with positive eigenvalues. Then, the Value at Risk is implicitly defined by:

$$\frac{1}{(\det(2\pi|D|))^{\frac{1}{2}}} \int_{\frac{1}{2}(|y^+|^2-|y^-|^2)\leq -VaR_p+\frac{1}{2}c'D^{-1}c} \exp\left(-\frac{1}{2}\langle y-f, |D|^{-1}(y-f)\rangle\right) dy = 1-p$$

where $f = |D|^{\frac{1}{2}}D^{-1}c$.

Now, the methodology consists in finding the asymptotic behavior of the function $\rho \mapsto I(\rho)$ defined by:

[29] In general, the reader should be careful when using the results of [16] since some of them were corrected by one of the authors in [17]. However, the methodology is of great interest since it generalizes to non-normal variables (see below and [17, 55]).

$$I(\rho) = \frac{1}{(\det(2\pi|D|))^{\frac{1}{2}}} \int_{\frac{1}{2}(|y^+|^2-|y^-|^2)\leq -\rho^2} \exp\left(-\frac{1}{2}\langle y-f, |D|^{-1}(y-f)\rangle\right) dy$$

Since $\lim_{\rho\to+\infty} I(\rho) = 0$ and $1-p$ is rather small, we can use an asymptotic equivalent $\tilde{I}(\rho)$ of $I(\rho)$ to find the solution ρ^* of the equation $I(\rho) = 1-p$.

This leads us to introduce $\tilde{\rho}^*$ implicitly given by $\tilde{I}(\tilde{\rho}^*) = 1-p$ and then the Value at Risk will be estimated by:

$$\widehat{VaR}_{t,t+1}(p) = \tilde{\rho}^{*2} - \frac{1}{2}c'D^{-1}c$$

Then, we see that the methodology boils down to the finding of a function $\tilde{I}$, easily invertible, such that $\lim_{\rho\to+\infty} \frac{I(\rho)}{\tilde{I}(\rho)} = 1$.
In the particular case of a Δ-hedged portfolio (which is often the case for portfolios of options), the approximation proposed in [17, 55] can be written as:

$$\begin{aligned}
\tilde{I}(\rho) &= \frac{1}{\sqrt{\pi}} \frac{(d_1^-)^{\frac{n-1}{2}}}{\sqrt{\prod_{j=2}^{n^-}(d_1^- - d_j^-)\prod_{j=1}^{n^+}(d_1^- + d_j^+)}} \int_{\frac{\rho^2}{d_1^-}}^{\infty} e^{-s} s^{-\frac{1}{2}} ds \\
&= \frac{1}{\sqrt{\pi}} \frac{(d_1^-)^{\frac{n-1}{2}}}{\sqrt{\prod_{j=2}^{n^-}(d_1^- - d_j^-)\prod_{j=1}^{n^+}(d_1^- + d_j^+)}} \Gamma\left(\frac{1}{2}, \frac{\rho^2}{d_1^-}\right)
\end{aligned}$$

where $\Gamma(\cdot,\cdot)$ stands here for the incomplete gamma function. Although the incomplete gamma function is implemented in most software, another possible choice for $\tilde{I}(\cdot)$ could be to consider an asymptotic expansion of the incomplete gamma function and this gives another $\tilde{I}(\cdot)$ defined as:

$$\tilde{I}(\rho) = \frac{1}{\sqrt{\pi}} \frac{(d_1^-)^{\frac{n}{2}}}{\sqrt{\prod_{j=2}^{n^-}(d_1^- - d_j^-)\prod_{j=1}^{n^+}(d_1^- + d_j^+)}} \frac{1}{\rho} e^{-\frac{\rho^2}{d_1^-}}$$

The only thing we need is in fact to be able to find numerically a solution to the equation $\tilde{I}(\rho) = 1-p$.

Now, we see in these functional forms that the eigenvalue $-d_1^-$ plays a specific role. This is

because the smallest eigenvalue is in the leading term of the asymptotic expansion. For that reason, the approach and the associated estimate of the Value at Risk is called principle component Value at Risk.

Now, coming to the more general case of a portfolio that has not been hedged in Δ, the approach can be adapted and the methodology to do so is exposed in appendix A of [16] and relies as above on the approximation of integrals over quadrics[30]. The method can also be straightforwardly adapted to take account of a term in Θ.

This methodology also has the advantage that it provides bounds for the Value at Risk of a quadratic portfolio hedged in Δ and not only an estimate. Hence, although the principal component Value at Risk is an approximation, the approach provides a range in which the Value at Risk lies (see [17]).
In practice this method is very fast but the range in which the Value at Risk is expected to lie is sometimes quite large in the examples exhibited in [17].

5.8 Partial Monte Carlo simulations

Let us now come to the last method we develop here to estimate the Value at Risk in the Delta-Gamma-(Theta)-Normal framework.
This methodology consists in going back to the definition of the quadratic approximation and estimating empirically the distribution using draws of the random variable at stake.

Recalling indeed that $\tilde{R} = r'_{t,t+1}\tilde{\Delta} + \frac{1}{2}r'_{t,t+1}\tilde{\Gamma}r_{t,t+1}$ with $r_{t,t+1} \sim \mathcal{N}(0,\Sigma)$, we see that we just need to simulate a variable with distribution $\mathcal{N}(0,\Sigma)$ and this is straightforward using Cholesky decomposition. Once we have a large quantity of draws for $\tilde{R}$, it is then easy to estimate the quantile corresponding to the Value at Risk we are looking for.

As for any Monte Carlo simulation, the difficulty arises from the very low rate of convergence toward the limit. Since the convergence rate is in the square root of the number of draws,

[30]Some uncorrected errors lie in the paper and only the methodology should be considered.

the number of draws necessary to have an acceptable measure is very large and the approach turns out to be very time-consuming when the number of risk factors is large. However, this method is, as the Fourier transform method, not based on an approximation and converges asymptotically toward the real Value at Risk of the quadratic portfolio. In other words, there is a trade-off between accuracy and time and the desired accuracy can – at least theoretically – always be achieved (within the quadratic portfolio approximation framework).

The difference between this partial Monte-Carlo method and the full Monte-Carlo that we discuss in the next part is that, here, the value of the portfolio is not recomputed after each draw using pricers but rather approximated using a Taylor expansion involving the Greeks. Even though it is less time-consuming than the full Monte-Carlo approach, the computation of the Value at Risk (of the quadratic portfolio) needs far more computing time than a Fourier inversion (if we are to compare exact methods and not approximations).

6 Going beyond normality

The strength of the parametric methodologies presented above, in the context of the Delta approach or of the Delta-Gamma-(Theta) approach, was that we were able to provide ways to approximate the Value at Risk for the desired confidence level using closed form expressions or to calculate numerically the Value at Risk of the approximated portfolio using basic numerical tools (Fast Fourier Transform being the most important example). This strength, which makes the analytical methodologies very fast compared to Monte-Carlo simulations, is however compensated by an important limitation on the distribution of the risk factors. So far, we have indeed only considered gaussian distributions for these risk factors, though conditional on the variance process.

Parametric methods require in fact two assumptions whose combination is of the utmost importance. First, a joint distribution is chosen for the risk factors. Second, an approximation is made for the return of the portfolios. Then, the resulting distribution for the approximated portfolio return must have properties that allow to approximate or calculate easily its quantiles.

In the case of the linear approximation of the portfolio return (Delta approach) – and obviously

also when the Theta term is taken into account –, the use of multivariate normal distributions allowed us to write the approximated Value at Risk in closed-form.
A general framework can in fact be developed to implicitly characterize the Value at Risk when the joint distribution of the risk factors' returns is elliptical or even a mixture of elliptical distributions. Although the formulae to characterize Value at Risk in this context are very straightforward to derive, a Delta-(Mixture) Elliptical approximation was proposed only a few years ago in [56].

In the more relevant case of the quadratic Delta-Gamma-(Theta) approximation for the returns of the portfolio, a joint normal distribution was used because the resulting distribution for the approximated portfolio had moments that were known in closed-form (this was used in the above paragraphs on Gaussian approximation, on the use of Johnson transformation, or on the Cornish-Fisher expansion[31]) and had a moment generating function / characteristic function that could be calculated using linear algebra (this was used in the Fourier transform approach and for the saddle-point approximation).

To use the methodologies relying on moments outside the gaussian framework, the important point to notice is that the entire moment generating function or the characteristic function can be computed explicitly, in the Delta-Gamma-(Theta) framework, for some non-gaussian distributions. We will indeed see that the methods based on these functions are perfectly suited to mixtures of gaussian distributions (see [34]) or when the risk factors follow some jump-diffusion processes (see [27]). Surprisingly also, these methods can be used when the returns of the risk factors follow a multivariate Student distribution, although this distribution has no moment generating function (see [40]). Finally, the so-called principal component Value at Risk – which is an approximation of the Value at Risk – can be calculated when the distribution of the risk factors' return is a Generalized Laplace distribution (this family of distributions is made of elliptical distributions and contains both the multivariate Laplace distributions and the multivariate Gaussian distributions)

[31] The Solomon-Stephens approximation is also an example, although it is not clear that this method can be used in practice.

6.1 The Delta-(Mixture) Elliptical approach

In the context of the Delta approximation, the portfolio return is approximated by:

$$\tilde{R} = \sum_{i=1}^{n} \tilde{\Delta}^i r_{t,t+1}^i$$

In the preceding paragraphs, we supposed that the return of the risk factors followed a multivariate normal distribution. Here we consider that the probability distribution function of $r_{t,t+1} = (r_{t,t+1}^1, \ldots, r_{t,t+1}^n)'$ is of the form:

$$\frac{1}{\sqrt{\det(\Sigma)}} g\left(r'\Sigma^{-1}r\right)$$

for some function g such that $\int_{\mathbb{R}^n} g(|y|^2)dy = 1$.

In that case, the Value at Risk of the linear portfolio is characterized by:

$$1 - p = \mathbb{P}\left(\tilde{\Delta}' r_{t,t+1} \leq -\widehat{VaR}_{t,t+1}(p)\right)$$

and this equation can be written in the following form:

$$1 - p = \int_{y_n=-\infty}^{+\infty} \cdots \int_{y_2=-\infty}^{+\infty} \int_{y_1=-\infty}^{-\frac{\widehat{VaR}_{t,t+1}(p)}{\|T_\Sigma' \tilde{\Delta}\|}} g(y_1^2 + y_2^2 + \cdots + y_n^2) dy_1 \cdots dy_n$$

where $\Sigma = T_\Sigma T_\Sigma'$ is the Cholesky decomposition of Σ.

This can be simplified to:

$$1 - p = |\mathbb{S}^{n-2}| \int_{\rho=0}^{+\infty} \int_{y_1=-\infty}^{-\frac{\widehat{VaR}_{t,t+1}(p)}{\|T_\Sigma' \tilde{\Delta}\|}} \rho^{n-2} g(y_1^2 + \rho^2) dy_1 d\rho$$

or

$$1 - p = |\mathbb{S}^{n-2}| \int_{\rho=0}^{+\infty} \int_{y_1=\frac{\widehat{VaR}_{t,t+1}(p)}{\|T_\Sigma' \tilde{\Delta}\|}}^{+\infty} \rho^{n-2} g(y_1^2 + \rho^2) dy_1 d\rho$$

where $|\mathbb{S}^{n-2}| = \frac{2\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})}$ is[32] the surface area of the unit sphere in $\mathbb{R}^{n-1}$.

[32]Here Γ is the usual Γ function.

Hence, the $\widehat{VaR}_{t,t+1}(p)$ can be written as:

$$\widehat{VaR}_{t,t+1}(p) = y(p)\|T'_{\Sigma}\tilde{\Delta}\| = y(p)\sqrt{\tilde{\Delta}'\Sigma\tilde{\Delta}}$$

where $y(p)$ solves $H(y(p)) = 1 - p$ with

$$H(y) = |\mathbb{S}^{n-2}| \int_{\rho=0}^{+\infty} \int_{y_1=y}^{+\infty} \rho^{n-2} g(y_1^2 + \rho^2) dy_1 d\rho$$

or equivalently:

$$H(y) = \frac{|\mathbb{S}^{n-2}|}{2} \int_{z=y}^{+\infty} dz \int_{u=z^2}^{+\infty} (u - z^2)^{\frac{n-3}{2}} g(u) du$$

This expression for the estimated Value at Risk is in fact a generalization of the expression proposed in the Delta-Normal case. In this case indeed, the form of function g allows to reduce the expression of H to the cumulative distribution function of a standard normal random variable. In general, the function H needs to be calculated numerically before we find $y(p)$ and is not a priori tabulated as it was in the gaussian case.

This approach is developed in [55, 56] and applied to three elliptical distributions:

- Student distribution
- Generalized Student distribution
- Generalized Laplace distribution

The authors show[33] that the function H can be written in closed-form using special functions. In the case of a Generalized Student distribution, H is indeed the sum of hypergeometric functions and in the case of a Generalized Laplace distribution, H can be written in closed-form using the incomplete Gamma function.

What is also interesting in this approach is that it easily generalizes to mixtures of elliptical distributions.

[33] The author does not seem to be aware that for a multivariate Student distribution with degree of freedom ν (*i.e.* $g(u) = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{n}{2}}} \left(1 + \frac{u}{\nu}\right)^{-\frac{\nu+n}{2}}$) the result is simply:

$$\widehat{VaR}_{t,t+1}(p) = -q_{1-p}^{t_\nu}\sqrt{\tilde{\Delta}'\Sigma\tilde{\Delta}}$$

where $q_{1-p}^{t_\nu}$ is the $(1-p)$-quantile of a 1D Student distribution with degree of freedom ν.

Imagine indeed that $r_{t,t+1}$ is distributed as a mixture of elliptical distributions given by probabilities $(\beta_j)_j$, and for each j an elliptical distribution with a dependence structure $\Sigma = \Sigma_j$ and a probability distribution function characterized by $g = g_j$.
Then, using the above notation, we simply have that the estimation of the Value at Risk is implicitly given by:

$$1 - p = |\mathbb{S}^{n-2}| \sum_j \beta_j \int_{\rho=0}^{+\infty} \int_{y_1=\frac{\widehat{VaR}_{t,t+1}(p)}{\|T'_{\Sigma_j}\tilde{\Delta}\|}}^{+\infty} \rho^{n-2} g_j(y_1^2 + \rho^2) dy_1 d\rho$$

In particular, if the different elliptical distributions that make the mixture share the same dependence structure matrix Σ – the mixture being then on tail-fatness for instance –, then, as above we can write the estimation of the Value at Risk as:

$$\widehat{VaR}_{t,t+1}(p) = y(p)\sqrt{\tilde{\Delta}'\Sigma\tilde{\Delta}}$$

where $y(p)$ solves $H(y(p)) = 1 - p$ with

$$H(y) = \frac{|\mathbb{S}^{n-2}|}{2} \int_{u=y}^{+\infty} dz \int_{u=z^2}^{+\infty} (u - z^2)^{\frac{n-3}{2}} \sum_j \beta_j g_j(u) du$$

This framework can be generalized to risk factors with a non-zero mean return and can similarly embed a Theta term. However, when it comes to generalizing to quadratic portfolios, other methods have to be developed.

In what follows, we start with the methods based on characteristic functions / moment generating functions. Then, we will present the general framework of the principal component Value at Risk. The latter approximation is indeed not restricted to gaussian returns but applies to the wider class of generalized Laplace distributions.

6.2 More on the saddle-point approximation and Fourier transform approaches

The saddle-point approximation and the Fourier tranform approach exposed in the context of the Delta-Gamma-Normal relied on a closed-form formula for the moment generating function or the characteristic function. Closed-form formulae can in fact be found in more general cases

and, as proposed by [34], this is the case of some gaussian mixtures where the risk factors' returns $r_{t,t+1}$ take the form of a random variable

$$r_{t,t+1} = \eta \overline{r}_{t,t+1}$$

where $\overline{r}_{t,t+1}$ follows a multivariate normal variable as in the preceding section for $r_{t,t+1}$ itself and where η is a (positive) scale random variable whose probability distribution function is a function $h(\cdot)$.

In that case, instead of having a decomposition of the form $\tilde{R} = \sum_j c_j Z_j + \frac{1}{2} d_j Z_j^2$ with $Z \sim \mathcal{N}(0, I)$ for the return of the quadratic portfolio as above, we simply have (because the decomposition only uses linear algebra):

$$\tilde{R} = \sum_j c_j \eta Z_j + \frac{1}{2} d_j \eta^2 Z_j^2$$

with $Z \sim \mathcal{N}(0, I)$ and for the same families of constants $(c_j)_j$ and $(d_j)_j$ as in the previous section.

Hence, after basic manipulations, the moment generating function can be written as:

$$M(t) = \int_0^\infty \overline{M}(t, u) h(u) du$$

where

$$\overline{M}(t, u) = \frac{1}{\sqrt{\prod_j (1 - tu^2 d_j)}} \exp\left(\frac{1}{2} \sum_j \frac{t^2 u^2 c_j^2}{1 - tu^2 d_j} \right)$$

Consequently, knowing h, we can numerically find M and, if it exists[34] outside of 0, use either the saddle-point approximation or a Fourier inversion to end up with an estimate of the Value at Risk at the desired confidence level.

The same applies for a finite mixture of gaussian variables. If indeed we consider a mixture of gaussians $\mathcal{N}(0, \Sigma_i)$ with weights β_i then the moment generating function is straightforwardly

[34] If it does not exist outside of 0, as it would be the case if $\frac{1}{\eta^2}$ is for instance a χ^2 variable – in that case the returns follow a multivariate Student distribution –, then a change of variables may apply to end up with the desired Value at Risk (see below).

given by:

$$M(t) = \sum_i \beta_i \frac{1}{\sqrt{\det(I - t\Sigma_i\tilde{\Gamma})}} \exp\left(\frac{1}{2}t^2\tilde{\Delta}'(\Sigma_i^{-1} - t\tilde{\Gamma})^{-1}\tilde{\Delta}\right)$$

Another framework, proposed in [27], in which one can compute the moment generating function is when the returns follow a multivariate jump-diffusion process with gaussian jumps. This framework corresponds to returns of the form:

$$r_{t,t+1} = X_0 - \lambda\mu + \sum_{i=1}^{N}(\mu + X_i)$$

where X_0 and the X_i's are independent multivariate normal variables with $X_0 \sim \mathcal{N}(0, \Sigma)$ and, for $i \geq 1$, $X_i \sim \mathcal{N}(0, V)$. The number of jumps N here is a Poisson process with intensity λ, independent of $(X_0, (X_i)_i)$.

As above, we will manage to find the moment generating function of

$$\tilde{R} = r'_{t,t+1}\tilde{\Delta} + \frac{1}{2}r'_{t,t+1}\tilde{\Gamma}r_{t,t+1}$$

because, conditionally on some other variable – here the number of jumps –, the distribution of $r_{t,t+1}$ is gaussian.

More precisely, we know that the moment generating function of $\tilde{R}$ is:

$$M(t) = \sum_{n=0}^{\infty} \mathbb{P}(N = n)\mathbb{E}\left[e^{tY_n}\right]$$

where

$$Y_n = w'_n\tilde{\Delta} + \frac{1}{2}w'_n\tilde{\Gamma}w_n$$

with $w_n \sim \mathcal{N}((n - \lambda)\mu, \Sigma + nV)$

Using what we have done in the previous section, we easily see that the moment generating function $t \mapsto M_n(t) = \mathbb{E}\left[e^{tY_n}\right]$ of Y_n is:

$$M_n(t) = \frac{1}{\sqrt{\det(I - t(\Sigma + nV)\tilde{\Gamma})}} \exp\left((n - \lambda)\mu' + \frac{1}{2}(n - \lambda)^2\mu'\tilde{\Gamma}\mu\right)$$

$$\times \exp\left(\frac{1}{2}t^2(\tilde{\Delta} + \tilde{\Gamma}(n - \lambda)\mu)'((\Sigma + nV)^{-1} - t\tilde{\Gamma})^{-1}(\tilde{\Delta} + \tilde{\Gamma}(n - \lambda)\mu)\right)$$

Hence, the moment generating function $M(\cdot)$ can be approximated using a finite number $\overline{n}$ of terms in the series defining it:

$$M(t) \simeq \sum_{n=0}^{\overline{n}} e^{-\lambda T} \frac{\lambda^n}{n!} M_n(t)$$

Then, the usual methodologies developed in the above section can be used to deduce the Value at Risk.

In the preceding examples, the possibility to extend the methods used in the Delta-Gamma-Normal framework was due to the random variable $\tilde{R}$ being normal, conditionally on some random variables: the scale random variable in the first example, the number of jumps in the last one. This can in fact be seen, between time t and $t+1$, as if there was a random volatility process.

A multivariate Student distribution may also appear as such a case. If indeed we recall that a multivariate Student distributed random variable may be represented as $\frac{(N_1,\ldots,N_n)}{\sqrt{\frac{Y}{\nu}}}$ where $(N_1,\ldots,N_n)' \sim \mathcal{N}(0,\Sigma)$ and $Y \sim \chi^2_\nu$, we see that conditionally on Y, the distribution is gaussian. However, although this is true, there is no moment generating function for a Student distribution.
To circumvent this problem, a trick is presented in [40] that more generally applies to the case where $r_{t,t+1}$ is of the form of our first example:

$$r_{t,t+1} = \eta \overline{r}_{t,t+1}$$

In that case, we have seen that the return of the quadratic portfolio is:

$$\tilde{R} = \sum_j c_j \eta Z_j + \frac{1}{2} d_j \eta^2 Z_j^2$$

with $Z \sim \mathcal{N}(0, I)$.

Instead of evaluating the moment generating function of $\tilde{R}$ that may not exist, we can remark

that the cumulative distribution function of $\tilde{R}$ we are eventually interested in is:

$$\begin{aligned} F(x) &= \mathbb{P}(\sum_j c_j \eta Z_j + \frac{1}{2} d_j \eta^2 Z_j^2 \leq x) \\ &= \mathbb{P}(\sum_j \frac{c_j}{\eta} Z_j + \frac{1}{2} d_j Z_j^2 - \frac{x}{\eta^2} \leq 0) = G_x(0) \end{aligned}$$

where G_x is the cumulative distribution function of

$$\sum_j \frac{c_j}{\eta} Z_j + \frac{1}{2} d_j Z_j^2 - \frac{x}{\eta^2}$$

Hence, we can focus on the moment generating function $\mathcal{M}$ of this random variable and we get, conditioning on η, that:

$$\mathcal{M}(t) = \mathbb{E}\left[\frac{1}{\sqrt{\prod_j (1 - t d_j)}} \exp\left(\frac{1}{2\eta^2} \sum_j \frac{t^2 c_j^2}{1 - t d_j} - \frac{tx}{\eta^2} \right) \right]$$

We see that, if the moment generating function $M_{\frac{1}{\eta^2}}$ of $\frac{1}{\eta^2}$ does exist outside of 0 – and this is the case in our multivariate Student case –, then the same is true for $\mathcal{M}$ and we have:

$$\mathcal{M}(t) = \frac{1}{\sqrt{\prod_j (1 - t d_j)}} M_{\frac{1}{\eta^2}}\left(\frac{1}{2} \sum_j \frac{t^2 c_j^2}{1 - t d_j} - tx \right)$$

In particular, when $\eta = \sqrt{\frac{Y}{\nu}}$ and the distribution of the risk factors' returns is a multivariate Student with degree of freedom ν we have:

$$\mathcal{M}(t) = \frac{1}{\sqrt{\prod_j (1 - t d_j)}} \left(1 - \frac{1}{\nu} \sum_j \frac{t^2 c_j^2}{1 - t d_j} + \frac{2tx}{\nu} \right)^{-\frac{\nu}{2}}$$

Once $\mathcal{M}$ is known, we can use a Fourier inversion to obtain G_x and the Value at Risk is solution of:

$$G_{-\widehat{VaR}_{t,t+1}(p)}(0) = 1 - p$$

We have now seen the main ideas to generalize the Fourier transform approach and the saddle-point approximation to non-gaussian variables. Another approach that can be adapted to

non-normal variables is the one based on principal component Value at Risk.

6.3 Principal component Value at Risk: a general framework

In a previous section, we presented the principal component Value at Risk in a gaussian framework. In reality, this approach is well suited, as exposed in [17], to a larger class of distributions. We will suppose indeed that, for a portfolio hedged in Δ:

$$\tilde{R} = \frac{1}{2} r'_{t,t+1} \tilde{\Gamma} r_{t,t+1}$$

where $r_{t,t+1}$ follows a Generalized Laplace distribution whose probability distribution function is:

$$f(x) = \frac{C_{\alpha,n}}{\sqrt{\det(\Sigma)}} \exp\left(-c_{\alpha,n}(x'\Sigma^{-1}x)^{\frac{\alpha}{2}}\right)$$

where:

$$C_{\alpha,n} = \frac{\alpha}{2\pi^{\frac{n}{2}}} \left(\frac{\Gamma(\frac{n+2}{\alpha})}{n\Gamma(\frac{n}{\alpha})}\right)^{\frac{n}{2}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{\alpha})}$$

and

$$c_{\alpha,n} = \left(\frac{\Gamma(\frac{n+2}{\alpha})}{n\Gamma(\frac{n}{\alpha})}\right)^{\frac{\alpha}{2}}$$

Doing the same diagonalization of $T'_\Sigma \tilde{\Gamma} T_\Sigma$ as above, where $\Sigma = T_\Sigma T'_\Sigma$, we obtain:

$$T'_\Sigma \tilde{\Gamma} T_\Sigma = \Omega D \Omega'$$

we suppose that the eigenvalues of D verify the same condition as in the gaussian case above (the smallest eigenvalue must be negative and simple):

$$-d_1^- < -d_2^- \leq \ldots \leq -d_{n^-}^- \leq 0 \leq d_1^+ \leq \ldots \leq d_{n^+}^+$$

Then, in this generalized context – for $\alpha = 2$ we are back to the gaussian case –, [16] introduced the Principal component Value at Risk as the solution of:

$$1 - p = A\Gamma\left(\frac{n}{\alpha} - \frac{n-1}{2}, c_{\alpha,n}\left(\frac{2\widehat{VaR}_{t,t+1}(p)}{d_1^-}\right)^{\frac{\alpha}{2}}\right)$$

with

$$A = 2\alpha^{-\frac{n+1}{2}}(2\pi)^{\frac{n-1}{2}} c_{\alpha,n}^{-\frac{n}{\alpha}} C_{\alpha,n} \frac{(d_1^-)^{\frac{n-1}{2}}}{\sqrt{\prod_{j=2}^{n^-}(d_1^- - d_j^-)\prod_{j=1}^{n^+}(d_1^- + d_j^+)}}$$

This approximation is valid asymptotically when $1-p$ tends to 0 and better when there is no cluster of eigenvalues near the smallest one, but the main strength of the approach is that it gives a range for where the Value at Risk of the quadratic portfolio lies.

7 Concluding remarks on the linear and quadratic approximations

The linear and the quadratic approximations have been shown to be quite useful to avoid numerous recalculations of the value of the portfolio. For most financial assets, Δs and Γs and other Greeks[35] are indeed available without additional computations than those usually done for trading. We have also seen that the combination of the linear and quadratic approximations with many distributional assumptions on the risk factors. Overall, for the analytical methodology to work, the risk factors' distribution has to be conditionally gaussian, although some other settings are possible as we have seen in the above section. Imposing conditionally gaussian models does not really reduce the set of possibilities since it incorporates GARCH models with normal residuals[36] and mixtures of gaussian distributions.

Now, if the linear approximation suffers from obvious limitations, noticeably for option portfolios that are often hedged in Δ, the quadratic approximation can induce good Value at Risk estimates for many portfolios. In most of the scientific articles dedicated to Value at Risk through a quadratic approximation of the portfolio value, examples are exhibited to compare the true Value at Risk and the estimates computed by the authors. The results, as expected, strongly depend on the nature of the portfolio and, for instance, it would be nonsense to use a quadratic approximation for a portfolio behaving as a call-spread. Some more complex portfolios can be found in [14] or [20] such that the quadratic approximation is misleading.
As a consequence, the very nature of the portfolio should be first analyzed before applying

[35] If volatility is one of the risk factors, the terms in Δ involve the dependence on volatility, or Vega.

[36] The initial EWMA RiskMetrics framework for the volatility being an instance of GARCH model. In [87], the RM 2006 framework is presented and is another GARCH model with long memory.

analytical methods to estimate Value at Risk.

Once the relevancy of the quadratic approximation has been checked, the different approaches above can be used. Chronologically, the first ones to have been used are methods based on moments (see the description above) and we believe that their development was due to limited computing capacity. Nowadays, more complex methods such as Fourier inversion can be developed and will not be time-consuming.

Today, the trade-off between accuracy and time is in favor of Fourier Inversion within the analytical tools and the very trade-off appears once we introduce Full Monte-Carlo simulations on the scene.

Full Monte Carlo simulations[37], which are going to be dealt with in the next part, are indeed more accurate than any method based on a portfolio approximation but they require multiple computations of portfolio value and are therefore really time-consuming. In the next part, we will see that, the quadratic approximation of this section, in addition to providing Value at Risk figures, helps designing variance reduction techniques for Monte-Carlo simulations.

[37]From now, Monte-Carlo simulations will always refer to Full Monte-Carlo simulations and not Partial Monte-Carlo simulation on an approximation of the portfolio.

Part III

Monte-Carlo simulations and stochastic algorithms

In the two preceding parts we have seen rather fast techniques to obtain an estimate of the Value at Risk of a portfolio. The first approaches developed, based on historical simulations were for most of them highly non-parametric and were therefore able to take into account the complex structure of dependence between the different risk factors at stake. However, relying on historical data has a negative counterpart: the lack of data. In the first part we indeed insisted on the fact that estimating an extreme quantile consists in using only a few hundred data and is therefore highly questionable, whatever the historical method used.
Contrary to these first approaches, analytical or parametric methods to compute Value at Risk only rely on historical data to fit some parameters (volatility for instance). Then they provide an estimation of the Value at Risk based on both an approximation of the portfolio and distributional assumptions on the risk factors. The limitations thus arise from the approximations that may be hardly adapted to extreme risks for certain portfolios but the main advantage is definitely the speed of the methods.

The third family of methodologies, which we present in this part, is based on Monte-Carlo simulations. These Monte-Carlo simulations use as many simulations as desired contrary to the historical approaches and do not suffer from the drawbacks of analytical methods. However, these methods turn out to be far more time-consuming.

8 Monte-Carlo simulations

Monte-Carlo approaches consist in simulating a large number of values for the risk factors (still assuming as in the analytical methods a certain distribution for them) and estimating the Value at Risk of the portfolio through the estimation of a quantile of the resulting distribution of P&L. Contrary to historical simulations, the number of simulations is not limited by historical data since we can draw as many trajectories as desired, the draws following a distribution calibrated

on historical data. However, the a priori unlimited number of draws is in practice constrained by the time-consuming process of reevaluating the portfolio after each draw. Since the portfolio value is not calculated through approximations, Monte-Carlo approaches need to embed, as it was the case for the historical approaches, the pricers used to price the different financial assets in the portfolio. However, due to the low convergence speed of Monte-Carlo simulations, the number of draws has to be large and the time spent in reevaluating the portfolio is the main issue in this third family of methods.

To make the above discussion more formal, the most simple Monte-Carlo approach consists in fitting a distribution f to risk factors or more often risk factors' returns (using past data) and drawing a large number M of new values for the risk factors, namely values of $(X_{t+1}^1, \ldots, X_{t+1}^n)$, that we refer to as:

$$(X_{t+1}^{1,m}, \ldots, X_{t+1}^{n,m}), \qquad 1 \leq m \leq M$$

Then, the time-consuming step of the Monte-Carlo simulation consists of the evaluation of the portfolio for these new values of the risk factors or equivalently the computation of the P&L associated to each draw:

$$\mathcal{P}_{t,t+1}^m = V(X_{t+1}^{1,m}, \ldots, X_{t+1}^{n,m}) - V(X_t^1, \ldots, X_t^n), \qquad 1 \leq m \leq M$$

Once this step has been carried out, the quantile estimate can be calculated using the appropriate quantile of the empirical distribution of the P&L, or using other techniques such as Extreme Value Theory. In the case of the empirical quantile, it corresponds to Monte-Carlo simulations in the usual sense of estimating a mean since we are interested in estimating quantities of the form $\mathbb{P}(\mathcal{P} \leq z) = \mathbb{E}[1_{\mathcal{P} \leq z}]$ where $\mathcal{P}$ is the random variable representing P&L.

Now, as for any Monte-Carlo approach, the main issue is to reduce the necessary number of simulations. These techniques are usually variance reduction techniques and we will present in the next section applications of Importance Sampling and stratification to cumulative distribution function estimation.

But, before going to these variance reduction techniques, we first present an idea developed by

Jamshidian and Zhu [52] called scenario simulations.
We noted above that the most time-consuming step in our Monte-Carlo simulation is by far the computation of the 1-period P&L $\mathcal{P}^m_{t,t+1}$ for each draw of the risk factors. Jamshidian and Zhu proposed to solve this issue to approximate the value of the portfolio using a grid. If indeed we have for each asset in the portfolio a (discrete) list of values corresponding to chosen values of the risk factors (this procedure being simplified by the fact that each asset depends only on a few risk factors), then in the Monte-Carlo simulation, the value of the portfolio can be interpolated from the precalculated values.
This approach is interesting as long as the necessary number of precomputations is low and far below the number of draws.

This idea is simple and, if the number of precomputations is large enough, it does not suffer from the drawbacks of the quadratic approximation developed in the above analytical methods. Also, other ideas have been developed to reduce the number of draws necessary to obtain a given precision. These variance reduction techniques are now being described and we chose to present the ones relying on the ideas developed in the preceding part – although their use is limited to portfolios that can be approximated through quadratic approximations. Other less technical Importance Sampling methods can obviously be used, using translations for instance.

9 Variance reduction techniques

When dealing with Monte-Carlo simulations, one of the basic challenges consists in reducing the number of draws necessary to obtain a given width of confidence interval. In other words, if we are to estimate a quantity $\mathbb{E}[1_{\mathcal{P}\leq z}]$, the challenge is to reduce the variance of the estimator. In our context of quantile estimation or cumulative distribution function computations in the tails, the naive estimators have large variance because the event under consideration is a rare event. The first idea that we present, for gaussian or conditionally gaussian risk factors, is an application of importance sampling which consists in changing the distribution of the risk factors, and hence of $\mathcal{P}$, so that the rare event becomes common.

Then, we will present a second idea called stratification and both these variance reduction techniques are going to borrow from the work done for quadratic portfolios in the part on

analytical methods (see [37, 39]).

9.1 Importance Sampling

Importance Sampling (often called IS) is a variance reduction technique that consists in a change of probability in order to artificially increase the probability of a rare event. In our case, the rare event at stake is the event $\{\mathcal{P} \leq z\}$ where z is a negative number in the left tail of the distribution of $\mathcal{P}$. Equivalently, and to stick to the notations introduced in the case of quadratic approximations for the value of portfolio, we are rather going to consider portfolio returns and hence the event $\{R \leq y\}$.

When the risk factors' returns are gaussian, we have seen in the above part that the portfolio's return R can be approximated through a quadratic function of the returns (see preceding part for the notations):

$$\tilde{R} = r'_{t,t+1}\tilde{\Delta} + \frac{1}{2}r'_{t,t+1}\tilde{\Gamma}r_{t,t+1}$$

As explained in the preceding part, this expression can be transformed using linear algebra and we obtain:

$$\tilde{R} = \sum_j c_j Z_j + \frac{1}{2}d_j Z_j^2 = Z'c + \frac{1}{2}Z'DZ$$

where $Z \sim \mathcal{N}(0, I)$ is obtained by linear combinations of the risk factors's returns and where c and D can be explicitly computed using linear algebra.

This approximation allowed us to find an estimate of the Value at Risk of the portfolio (in terms of return), and we are going to evaluate the true cumulative distribution function of R in the neighborhood of this value. Consider indeed $y = -v$ where v is in the neighborhood of the Value at Risk, then our goal is to evaluate $\mathbb{P}(R \leq y) = \mathbb{E}[1_{R \leq y}]$.

In this approach we are going to draw Z and reconstitute the returns of the risk factors using linear transformation. Our change of probability will therefore affect primarily the variable Z, and the basic idea is to transform Z into another gaussian variable $\mathcal{N}(\mu, B)$ (under the new

probability that we call $\mathbb{Q}$) such that the event $\{R \leq y\}$ or more exactly $\{\tilde{R} \leq y\}$ becomes common. The idea then is to choose $\mathbb{Q}$ such that:

$$\mathbb{E}^{\mathbb{Q}}\left[\tilde{R}\right] = \mathbb{E}\left[\tilde{R}\frac{|\det(B)|^{\frac{1}{2}}\exp\left(-\frac{1}{2}(Z-\mu)'B^{-1}(Z-\mu)\right)}{\exp\left(-\frac{1}{2}Z'Z\right)}\right] = y$$

To choose $\mathbb{Q}$, the authors in [37, 39] propose to consider:

$$B = (1-\theta D)^{-1}, \qquad \mu = \theta(1-\theta D)^{-1}c$$

where θ is the variable to be chosen in order to satisfy $\mathbb{E}^{\mathbb{Q}}\left[\tilde{R}\right] = y$.

In that case indeed, the above likelihood ratio can be expressed easily as:

$$\frac{|\det(B)|^{\frac{1}{2}}\exp\left(-\frac{1}{2}(Z-\mu)'B^{-1}(Z-\mu)\right)}{\exp\left(-\frac{1}{2}Z'Z\right)} = \exp(\theta\tilde{R} - K(\theta))$$

where $K(\cdot)$ is the cumulant generating function of $\tilde{R}$ defined as the logarithm of the moment generating function $M(\cdot)$.
Thus the choice of θ reduces to solving:

$$\mathbb{E}\left[\tilde{R}\exp(\theta\tilde{R} - K(\theta))\right] = y$$

i.e.

$$\frac{M'(\theta)}{M(\theta)} = K'(\theta) = y$$

This equation can be solved easily, at least numerically, since we know the moment generating function of $\tilde{R}$ in closed-form (see the preceding part).

Once θ is chosen, the problem consists in using a Monte-Carlo simulation on

$$\mathbb{E}\left[1_{R\leq y}\right] = \mathbb{E}^{\mathbb{Q}}\left[1_{R\leq y}\frac{\exp\left(-\frac{1}{2}Z'Z\right)}{|\det(B)|^{\frac{1}{2}}\exp\left(-\frac{1}{2}(Z-\mu)'B^{-1}(Z-\mu)\right)}\right]$$

where Z is simulated as a gaussian variable $\mathcal{N}(\mu, B)$ and where R is calculated as a function

of the risk factors, the risk factors being linearly recalculated from Z.

Now, the twisting factor θ may be chosen for one given y and fixed to compute other values of the cumulative distribution function than at point y. Several Monte-Carlo simulations need indeed be done, but fixing θ allows not to recompute the whole portfolio values and reduces the problem to the computation of different averages.

9.2 Stratification

Another methodology than IS to reduce variance is stratified sampling or stratification. This idea is quite general and requires once again an approximation of the variable under scrutiny – this approximation being in our case the same quadratic approximation as above.

This general idea is quite simple and consists, when one knows an auxiliary variable $\tilde{Y}$ (here $\tilde{R}$) correlated to the variable Y (here $1_{R>y}$) whose mean is to be estimated, in partitioning the space of the auxiliary variable in k strata denoted $A_1, \ldots, A_k$ and evaluating the mean under scrutiny using the formula

$$\mathbb{E}[Y] = \sum_{i=1}^{k} \mathbb{E}[Y|A_i]\mathbb{P}(A_i)$$

where each of the k conditional means is calculated by a Monte-Carlo simulation. Obviously, the success of stratification depends strongly on the number of points used for each of the k condition mean.

A simple example that provides variance reduction is to consider $M_i = \mathbb{P}(A_i)M$ simulations for the i^{th} conditional mean.

In practice[38], and many techniques have been proposed (see for instance [19], [37] or [39]), one can first stratify the space for $\tilde{R}$ and compute, using what we called Partial Monte-Carlo simulations, the probabilities $\mathbb{P}(A_i)$. Then, the method consists in drawing values of Z so as to constitute (dropping values when necessary) a set of M values with $M_i = \mathbb{P}(A_i)M$ values of $\tilde{R}$

[38]This stratification technique can obviously be used in addition to the above importance sampling approach.

in strata A_i for each index i. Then Full Monte-Carlo simulations with computation of the portfolio's return can be carried out on each strata to end up with the estimate of the Value at Risk.

Importance sampling and stratified sampling are useful to reduce variance in Monte-Carlo simulations. However, IS can be used in other contexts than pure Monte-Carlo simulation. We present in the next section another useful approach that enters that vast family of random simulations and consists in estimating directly the Value at Risk through a stochastic algorithm instead of building the cumulative distribution function as in the above Monte-Carlo simulations.

10 Stochastic algorithms

Monte-Carlo simulations are well suited to finding the cumulative distribution function of the 1-period P&L. However, since we are eventually interested in a quantile of this distribution, it may be interesting to understand how one can directly estimate the Value at Risk using the same simulated data as in a Monte-Carlo approach. In what follows, we are going to present an approach based on stochastic algorithms that allows us to directly estimate the Value at Risk. This approach, with quite technical proofs and extensions, is presented in [5].

10.1 The Bardou-Frikha-Pagès approach

The approach we present differs from Monte-Carlo in only a few ways. The main common point concerns simulated data. We indeed have to make distributional assumptions on the risk factors and then compute a collection of P&L values $(\mathcal{P}^m_{t,t+1})_m$ based on the M draws made for the risk factors' returns. However, this collection of P&L values are not used to build a cumulative distribution function but rather to progressively build a sequence of Value at Risk estimates. An important consequence is that the order in which points are taken into consideration has some importance. We guess that changing the order and running the algorithms for each new order of the data before averaging may provide better estimates.

In fact the approach relies on the fact that the true Value at Risk V at confidence level p

verifies:

$$\mathbb{E}[H(\mathcal{P}, V)] = 0, \qquad H(P, V) = 1 - \frac{1}{1-p} 1_{P \leq -V}$$

Hence, using Robbin-Monro theorem, the method to estimate V consists in building recursively a sequence $(V^m)_m$ of Value at Risk estimates by:

$$V^m = V^{m-1} - \gamma^m H\left(\mathcal{P}_{t,t+1}^m, V^{m-1}\right)$$

with V^0 fixed and $(\gamma^m)_m$ a deterministic positive sequence verifying $\sum_{m=0}^{\infty} \gamma^m = +\infty$ and $\sum_{m=0}^{\infty} (\gamma^m)^2 < +\infty$

Under mild assumptions, it is proven in [5] that the above sequence $(V^m)_m$ converges almost surely towards the true Value at Risk and the authors provide a confidence interval using the asymptotic normality of an estimator derived from $(V^m)_m$.

This algorithm, in its formulation, recalls in some sense the baseline CAViaR specification for quantile regression, although the underlying ideas are completely different. We indeed have that the Value at Risk estimate decreases slowly as long as no value $\mathcal{P}_{t,t+1}^m$ oversteps the most recent Value at Risk estimate and increases in important proportion in the opposite case.

As for the above Monte-Carlo simulations, it may be interesting to use Importance Sampling to make the rare event of exceeding the Value at Risk estimate a more common event. This is the purpose of the next paragraphs.

10.2 Importance sampling for stochastic algorithms

Importance sampling helped in the case of Monte-Carlo simulations in making a rare event more common through a change of probability. Here, because the rare event at step m consists in a P&L figure exceeding the current estimate V^{m-1} of the Value at Risk, importance sampling must be adaptative.

If we focus on Importance Sampling through a translation in the risk factors or more precisely

in the risk factors returns (more general Importance Sampling methods, as of the kind used above, can also be used and we recommend a thorough reading of [5] for details), our goal is to find $\theta^{m-1} \in \mathbb{R}^n$ such that the simulated data $\mathcal{P}^m_{\theta^{m-1},t,t+1}$ defined by

$$\mathcal{P}^m_{\theta^{m-1},t,t+1} = V(X^1_t(1+r^{1,m}_{t,t+1}+\theta^{1,m-1}),\dots,X^n_t(1+r^{n,m}_{t,t+1}+\theta^{n,m-1}))-V(X^1_t,\dots,X^n_t), \qquad 1 \le m \le M$$

has a higher probability of exceeding the current threshold V^{m-1} than the usual simulated data:

$$\mathcal{P}^m_{t,t+1} = V(X^{1,m}_{t+1},\dots,X^{n,m}_{t+1}) - V(X^1_s,\dots,X^n_s)$$

$$= V(X^1_t(1+r^{1,m}_{t,t+1}),\dots,X^n_t(1+r^{n,m}_{t,t+1})) - V(X^1_s,\dots,X^n_s), \qquad 1 \le m \le M$$

The authors provide an interpretation of the best translation parameter θ as the solution of an equation for which a stochastic gradient descent can be applied and they eventually propose the following adaptative importance sampling:

$$V^m = V^{m-1} - \gamma^m e^{-\rho|\theta^{m-1}|^b}\left(1 - \frac{1}{1-p}1_{\mathcal{P}^m_{\theta^{m-1},t,t+1}\le -V}\frac{f(r^{m-1}_{t,t+1}+\theta^{m-1})}{f(r^{m-1}_{t,t+1})}\right)$$

$$\theta^m = \theta^{m-1} - \gamma^m e^{-2\rho|\theta^{m-1}|^b}1_{\mathcal{P}^m_{-\theta^{m-1},t,t+1}\le -V}\frac{f(r^{m-1}_{t,t+1}-\theta^{m-1})^2}{f(r^{m-1}_{t,t+1})f(r^{m-1}_{t,t+1}-2\theta^{m-1})}\frac{\nabla f(r^{m-1}_{t,t+1}-2\theta^{m-1})}{f(r^{m-1}_{t,t+1}-2\theta^{m-1})}$$

where f is the distribution assumed for the risk factors' returns which is supposed to satisfy technical conditions[39] that *define* b and ρ, namely:

$$\exists b \in [1,2], \frac{|\nabla f(z)|}{f(z)} = \mathcal{O}(|z|^{b-1})$$

$$\exists \rho > 0, \log(f(z)) + \rho|z|^b \text{ is convex}$$

The authors tested this IS procedure and some more complex ones and found them quite efficient. This literature being quite new, there is however, to our knowledge, no comparison with the usual Monte-Carlo simulations that relies on the same data.

[39] These conditions are satisfied under gaussian assumptions.

Part IV

From theory to practice

11 Meeting the practitioners

To complement the above review of the academic literature and to understand the needs of practitioners, we decided to carry out interviews of Chief Risk Officers and their team in a panel of banks and insurance companies and to meet people that are auditing their models in France at the ACP (Autorité de Contrôle Prudentiel). Most banks use Value at Risk for regulatory reasons (particularly to compute the regulatory capital) and as an internal risk measure. Because of the well-known drawbacks of Value at Risk, many banks complement VaR figures with Expected Shortfall indices, but Value at Risk seems to remain central for internal risk measurement and management. Most banks use indeed VaR figures to decide on front office limits and this fact has important consequences on the methodology – this methodology being often the same for management and regulatory purposes.
It appeared from our discussion with the practitioners that one of the main advantages of Value at Risk is that it is easy to communicate on. Similarly, VaR figures are often required to be hardly challengeable, neither by the top management, nor by the traders, and the methodologies to compute them must be widely accepted and understood inside each company for VaR to be an effective risk management tool.

In companies where this requirement on Value at Risk measures is predominant, historical simulations are used to compute VaR figures, most often over 1 day and then multiplied by $\sqrt{10}$. Historical simulations have indeed the advantage to be easily understandable and based on few parametric hypotheses, if any. This pragmatical reason is often complemented by a deeper one, more theoretical, regarding the interdependence structure of the risk factors. Historical simulations are often used because, portfolios being too complex and the number of risk factors being too large (sometimes over 1000), estimating the joint distribution of risk factors – be it assumed multivariate normal or not – will end up to doubtful figures and specifications.
These historical simulations are often carried out over a dataset of two years (around 500 days) and the Value at Risk is determined using the resulting "empirical" distribution of PnL, picking

the 5^{th} worst case for a 99% VaR or sometimes interpolating between two successive points to obtain a better estimate. Some banks in our sample also used bootstrapping techniques to improve their quantile estimates and Hill estimator (see the section on Extreme Value Theory) appeared to be also used. It seems however that these tools to potentially improve quantile estimates could be used in a more systematic way since their implementation only requires minor changes to the processes already in use: a Jackknife estimator or a tail index estimator could indeed be used after the usual historical simulation has been done and the "empirical" distribution obtained.

Also, concerning the academic improvements presented in Part 1, practitioners seem to be unwilling to overweight the present, as proposed in the BRW approach, or to take account of the current level of volatility (as in the Hull and White proposal). The main argument underlying this choice is, to our understanding, that it increases the procyclical dimension of Value at Risk. A consequence of that procyclicality is that it increases systemic risk and hence the regulator may not be keen to validate models including these features. Although they certainly provide a better estimate of the current VaR, BRW and Hull and White approaches are perhaps reflecting too much an intrinsic drawback of the Value at Risk to be good measures from a regulatory point of view.

Now, another reason why historical simulations are favored in a large part of the industry is because other methods may be too complex and would result in time-consuming processes incompatible with the daily delivery of a VaR figure. Banks using Monte-Carlo simulations on complex portfolios seem to have invested indeed a lot more in hardware and parallel computing than those using historical simulations.

If some banks are using historical simulations, other banks favor Monte-Carlo simulations and banks are clearly divided into two categories: those using historical simulations for the reasons explained above and those that are against the use of historical simulations because of the hypotheses underlying them (see Part 1). Some banks are using parametric methods to compute Value at Risk but these VaR figures are not used on a standalone basis and seem to be used on not too complex portfolios, in addition to other VaR figures, based on Monte-Carlo or historical simulations. The Delta-Gamma-Theta-Vega approaches are criticized because they approximate non-linearities locally whereas Value at Risk is linked to the tails of the distribution, and because they are not suited to complex payoffs. A natural consequence is that the

variance reduction techniques presented in Part 3 of this document and relying on parametric methods seem not to be used in practice.
Coming to the banks favoring Monte-Carlo simulations, the motivations are based on the theoretical drawbacks of historical simulations. However, while relying on Monte Carlo simulations, the advantages of historical simulations disappeared: specifying the dependence between risk factors is a complex process and the resulting joint distribution, which is often (log)-normal, is subject to criticisms. However, going beyond normal assumptions seem to be regarded as useless in practice, the calibration problem being of far greater importance. Regarding the convergence speed of algorithms, people favoring Monte-Carlo simulations often argue that since simulations are slow to converge, historical simulations with 2 years of data cannot provide good estimates. Proprietary models and methods are often used to compute Value at Risk with Monte-Carlo methods and it seems that an important investment has been made in parallel computing and sometimes toward the use of graphics processing units (GPU). Somehow, hardware developments seem to come first, before variance reduction techniques. Finally, the stochastic algorithms that could replace Monte-Carlo techniques seem not to be used in banks, certainly because it is too recent an approach and because models cannot be changed or even modified easily, both for technical and regulatory reasons.

While interviewing the risk teams, it also appeared that very important problems in Value at Risk computations are not tackled or even discussed in the academic literature. The choice of the risk factors is indeed of utmost importance in practice and is seldom discussed. Furthermore, this choice is twofold. New risks appear that were not taken into account and this requires changes in the model. Also, the right nature of the shocks to take into account is a real issue. When should a shock be considered additive or multiplicative? This question, that is particularly relevant in fixed income and for credit spreads, has no clear answer but important consequences. Another problem that is seldom tackled in the literature is the one linked to the time horizon of the Value at Risk. 10-day VaR or 1-year VaR must indeed rely on overlapping data, be it in the calibration step of Monte-Carlo simulations or in the construction of the historical sample for historical simulations. Another direction in which practitioners would like research to be done is to understand the individual contributions of each asset to Value at Risk.

12 Concluding remarks and perspectives

Our discussions with practitioners from various horizons – in banks or insurance companies, both French and international, or in charge of the regulation – allowed us to shed light on common practices. First, methods coming from statistics such as bootstrapping or EVT are rarely used and could be implemented easily on top of what is used today. Similarly, it seems that variance reduction techniques are not systematically used by companies who choose to rely on Monte-Carlo simulations. This is however difficult to adapt since variance reduction techniques should be coded within the algorithms and not on top of them. Another important point of our discussions is that academics are often developing methods that are not scalable to real data (with sometimes more than 1000 risk factors) or difficult to apply in practice, non-gaussian risk factors being the main examples.

Now, changes in the models used by banks, that consist in practice in complex processes, must be validated by the regulators and VaR models are therefore subject to an important agency problem in which the incentives of the regulator and the financial industry are not always the same. Models are often modified when too many exceptions are observed and if the current high volatility of the market comes to an end, exceptions are not going to be that numerous in the future, resulting in few modifications of the models, increasing therefore the capital requirement.

References

[1] C. Albanese, K. Jackson, and P. Wiberg. A new fourier transform algorithm for value-at-risk. *Quantitative Finance*, 4(3):328–338, 2004.

[2] M. Ammann and C. Reich. Var for nonlinear financial instruments-linear approximation or full monte carlo? *Financial Markets and Portfolio Management*, 15(3):363–378, 2001.

[3] F. Audrino and G. Barone-Adesi. A multivariate fgd technique to improve var computation in equity markets. *Computational Management Science*, 2(2):87–106, 2005.

[4] A.A. Balkema and L. De Haan. Residual life time at great age. *The Annals of Probability*, 2(5):792–804, 1974.

[5] O. Bardou, N. Frikha, and G. Pages. Computing var and cvar using stochastic approximation and adaptive unconstrained importance sampling. *Monte Carlo Methods and Applications*, 15(3):173–210, 2009.

[6] O. Bardou, N. Frikha, and G. Pagès. Recursive computation of value-at-risk and conditional value-at-risk using mc and qmc. *Monte Carlo and Quasi-Monte Carlo Methods 2008*, pages 193–208, 2009.

[7] G. Barone-Adesi and K. Giannopoulos. Non parametric var techniques. myths and realities. *Economic Notes*, 30(2):167–181, 2001.

[8] G. Barone-Adesi, K. Giannopoulos, and L. Vosper. Var without correlations for non-linear portfolios. *Journal of Futures Markets*, 19(5):583–602, 1999.

[9] T.S. Beder. Var: Seductive but dangerous. *Financial Analysts Journal*, pages 12–24, 1995.

[10] J. Berkowitz and J. O'Brien. How accurate are value-at-risk models at commercial banks? *The Journal of Finance*, 57(3):1093–1111, 2002.

[11] M. Billio and L. Pelizzon. Value-at-risk: a multivariate switching regime approach. *Journal of Empirical Finance*, 7(5):531–554, 2000.

[12] G. Bormetti, V. Cazzola, G. Livan, G. Montagna, and O. Nicrosini. A generalized fourier transform approach to risk measures. *Journal of Statistical Mechanics: Theory and Experiment*, 2010:P01005, 2010.

[13] J. Boudoukh, M. Richardson, and R. Whitelaw. The best of both worlds. *Risk*, 11(5):64–67, 1998.

[14] M. Britten-Jones and S.M. Schaefer. Non-linear value-at-risk. *European Finance Review*, 2(2):161, 1999.

[15] S.A. Broda. The expected shortfall of quadratic portfolios with heavy-tailed risk factors. *Mathematical Finance*.

[16] R. Brummelhuis, A. Cordoba, M. Quintanilla, and L. Seco. Principal component value at risk. *Mathematical Finance*, 12(1):23–43, 2002.

[17] R. Brummelhuis and J. Sadefo Kamdem. Var for quadratic portfolio's with generalized laplace distributed returns. *University of Reims working paper*, 2004.

[18] J. Cardenas, E. Fruchard, E. Koehler, C. Michel, and I. Thomazeau. Var: one step beyond. *Risk*, 10(10):72–75, 1997.

[19] J. Cardenas, E. Fruchard, J.F. Picron, C. Reyes, K. Walters, and W. Yang. Monte carlo within a day. *Risk*, 12(2):55–59, 1999.

[20] G. Castellacci and M.J. Siclari. The practice of delta-gamma var: Implementing the quadratic portfolio model. *European Journal of Operational Research*, 150(3):529–545, 2003.

[21] Rongda Chen, Dan Cao, and Qingyang Yu. Nonlinear var model of options portfolio under multivariate mixture of normals distributions. *Business Intelligence and Financial Engineering, International Conference on*, 0:364–368, 2010.

[22] V. Chernozhukov and L. Umantsev. Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics*, 26(1):271–292, 2001.

[23] P. Christoffersen. Value–at–risk models. *Handbook of Financial Time Series*, pages 753–766, 2009.

[24] M. Crouhy, D. Galai, and R. Mark. A comparative analysis of current credit risk models* 1. *Journal of Banking & Finance*, 24(1-2):59–117, 2000.

[25] L. De Haan and L. Peng. Comparison of tail index estimators. *Statistica Neerlandica*, 52(1):60–70, 1998.

[26] D. Duffie and J. Pan. An overview of value at risk. *The Journal of derivatives*, 4(3):7–49, 1997.

[27] D. Duffie and J. Pan. Analytical value-at-risk with jumps and credit risk. *Finance and Stochastics*, 5(2):155–180, 2001.

[28] L. El-Jahel, W. Perraudin, and P. Sellin. Value at risk for derivatives. *The Journal of Derivatives*, 6(3):7–26, 1999.

[29] P. Embrechts. Extreme value theory in finance and insurance. *Manuscript, Department of Mathematics, ETH, Swiss Federal Technical University*, 1999.

[30] P. Embrechts. Extreme value theory: potential and limitations as an integrated risk management tool. *Manuscript, Department of Mathematics, ETH, Swiss Federal Technical University*, 1999.

[31] R.F. Engle and S. Manganelli. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, pages 367–381, 2004.

[32] W. Fallon, Wharton School, and W.F.I. Center. *Calculating value-at-risk*. Citeseer, 1996.

[33] J. Fan and J. Gu. Semiparametric estimation of value at risk. *Econometrics Journal*, 6(2):261–290, 2003.

[34] A. Feuerverger and A.C.M. Wong. Computation of value-at-risk for nonlinear portfolios. *Journal of Risk*, 3:37–56, 2000.

[35] R.A. Fisher and L.H.C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press, 1928.

[36] J. Gil-Pelaez. Note on the inversion theorem. *Biometrika*, 38(3-4):481, 1951.

[37] P. Glasserman, P. Heidelberger, and P. Shahabuddin. *Importance sampling and stratification for value-at-risk*. Citeseer, 1999.

[38] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Efficient monte carlo methods for value-at-risk. 2000.

[39] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Variance reduction techniques for estimating value-at-risk. *Management Science*, pages 1349–1364, 2000.

[40] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, 12(3):239–269, 2002.

[41] BV Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *The Annals of Mathematics*, 44(3):423–453, 1943.

[42] C. Gourieroux and J. Jasiak. *Truncated maximum likelihood, goodness of fit tests and tail analysis*. Citeseer, 1998.

[43] M. Guidolin and A. Timmermann. Value at risk and expected shortfall under regime switching. *University of California at San Diego working paper*, 2003.

[44] E.J. Gumbel. *Statistics of extremes*. Dover Pubns, 2004.

[45] D. Hendricks. Evaluation of value-at-risk models using historical data. *Economic Policy Review*, 2(1):39–69, 1996.

[46] B.M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.

[47] ID Hill, R. Hill, and RL Holder. Fitting johnson curves by moments. *Applied Statistics*, 25(2):180–189, 1976.

[48] J.R.M. Hosking and J.R. Wallis. Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3):339–349, 1987.

[49] L.K. Hotta, E.C. Lucas, and H.P. Palaro. Estimation of var using copula and extreme value theory. *Multinational Finance Journal*, 12(3/4):205–218, 2008.

[50] J. Hull and A. White. Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk*, 1(1):5–19, 1998.

[51] J. Hull and A. White. Value at risk when daily changes in market variables are not normally distributed. *Journal of derivatives*, 5:9–19, 1998.

[52] F. Jamshidian and Y. Zhu. Scenario simulation: Theory and methodology. *Finance and Stochastics*, 1(1):43–67, 1996.

[53] S.R. Jaschke. The cornish-fisher expansion in the context of delta-gamma-normal approximations. *Journal of Risk*, 4:33–52, 2002.

[54] T. Javanainen. Analytical delta-gamma var methods for portfolios of electricity derivatives. 2004.

[55] J.S. Kamdem. Méthodes analytiques pour le risque des portefeuilles financiers. 2004.

[56] J.S. Kamdem. Var and es for linear portfolios with mixture of elliptic distributions risk factors. *Computing and Visualization in Science*, 10(4):197–210, 2007.

[57] R. Kawata and M. Kijima. Value-at-risk in a market subject to regime switching. *Quantitative Finance*, 7(6):609–619, 2007.

[58] R. Koenker and G. Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

[59] K. Kuester, S. Mittnik, and M.S. Paolella. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4(1):53, 2006.

[60] G. Lambadiaris, L. Papadopoulou, G. Skiadopoulos, and Y. Zoulis. Var: history or simulation? *Risk*, 16(9):122–127, 2003.

[61] Y. Lan, Z.M. Hu, and J. Johnson. Improving accuracy and precision of value-at-risk forecasts. 2010.

[62] A. Lehar. Alternative value-at-risk models for options. *Computing in Economics and Finance 2000*, 2000.

[63] R. Lugannani and S. Rice. Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, pages 475–490, 1980.

[64] C. Marshall and M. Siegel. Value at risk: Implementing a risk measurement standard. *Journal of Derivatives*, 4(3):91–110, 1997.

[65] A.J. McNeil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3-4):271–300, 2000.

[66] D. Meneguzzo, W. Vecchiato, and B. Intesa. Improvement on value at risk measures by combining conditional autoregressive and extreme value approaches. *Review in International Business and Finance*, pages 275–324, 2000.

[67] J.C. Meyfredi. Towards the integration of extreme values in obtaining the value-at-risk. *EDHEC Business School*, 2003.

[68] J. Mina and A. Ulmer. Delta-gamma four ways. *New York: RiskMetrics Group, Working Paper*, 1999.

[69] J. Mina, J.Y. Xiao, et al. Return to riskmetrics: the evolution of a standard. *RiskMetrics Group*, 2001.

[70] JP Morgan and Reuters. *Riskmetrics: technical document*. Morgan Guaranty Trust Company of New York, 1996.

[71] S. Pichler and K. Selitsch. A comparison of analytical var methodologies for portfolios that include options. *Model risk, concepts, calibration, and pricing*, 2000.

[72] J. Pickands. Statistical inference using extreme order statistics. *the Annals of Statistics*, 3(1):119–131, 1975.

[73] G. Piroozfar. Forecasting value at risk with historical and filtered historical simulation methods. *UUDM Project Report*, 15, 2009.

[74] M. Pritsker. Evaluating value at risk methodologies: accuracy versus computational time. *Journal of Financial Services Research*, 12(2):201–242, 1997.

[75] M. Pritsker. The hidden dangers of historical simulation. *Journal of Banking & Finance*, 30(2):561–582, 2006.

[76] C. Rouvinez. Going greek with var. *Risk*, 10(2):57–65, 1997.

[77] J. Sadefo Kamdem. Value-at-risk and expected shortfall for quadratic portfolio of securities with mixture of elliptic distributed risk factors. 2003.

[78] H. Solomon and Stephens M. A. Distribution of a sum of weighted chi-square variables. *Technical report*, 1977.

[79] J.W. Taylor. Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics*, 6(3):382, 2008.

[80] R.W.J. van den Goorbergh, P.J.G. Vlaar, and N. Bank. *Value-at-Risk Analysis of Stock Returns Historical Simulation, Variance Techniques or Tail Index Estimation?* Citeseer, 1999.

[81] S. Venkataraman. Value at risk for a mixture of normal distributions: the use of quasi-bayesian estimation techniques. *Economic Perspectives - Federal Reserve Bank of Chicago*, 21:2–13, 1997.

[82] P. Wiberg. Computation of value-at-risk: the fast convolution method, dimension reduction and perturbation theory. 2002.

[83] M.L. Yueh and M.C.W. Wong. Analytical var and expected shortfall for quadratic portfolios. *The Journal of Derivatives*, 17(3):33–44, 2010.

[84] P. Zangari. An improved methodology for measuring var. *RiskMetrics Monitor*, 2:7–25, 1996.

[85] P. Zangari. A var methodology for portfolios that include options. *RiskMetrics Monitor*, 1:4–12, 1996.

[86] G. Zumbach. The riskmetrics 2006 methodology. *RiskMetrics Group (www. risk metrics. com)*, 2006.

[87] G. Zumbach. A gentle introduction to the rm 2006 methodology. *New York, Riskmetrics Group*, 2007.