Impact of multimodality of distributions on VaR and ES calculations

Dominique Guegan, Bertrand Hassani, Kehan Li

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers

Impact of multimodality of distributions on VaR and ES calculations

Dominique GUEGAN, Bertrand HASSANI, Kehan LI

2017.19

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://centredeconomiesorbonne.univ-paris1.fr/
ISSN : 1955-611X

Impact of multimodality of distributions on VaR and ES calculations

Dominique Guégan[a], Bertrand Hassani[b], Kehan Li[c]

[a] *Université Paris 1 Panthéon-Sorbonne, CES UMR 8174. 106 bd l'Hopital 75013, Paris, France. Labex ReFi, Paris France. IPAG, Paris, France.*
[b] *Grupo Santander and Université Paris 1 Panthéon-Sorbonne, CES UMR 8174. Labex ReFi.*
[c] *Université Paris 1 Panthéon-Sorbonne, CES UMR 8174. 106 bd l'Hopital 75013, Paris, France. Labex ReFi.*

Abstract

Unimodal probability distribution has been widely used for Value-at-Risk (VaR) computation by investors, risk managers and regulators. However, financial data may be characterized by distributions having more than one modes. Using a unimodal distribution may lead to bias for risk measure computation. In this paper, we discuss the influence of using multimodal distributions on VaR and Expected Shortfall (ES) calculation. Two multimodal distribution families are considered: Cobb's family and distortion family. We provide two ways to compute the VaR and the ES for them: an adapted rejection sampling technique for Cobb's family and an inversion approach for distortion family. For empirical study, two data sets are considered: a daily data set concerning operational risk and a three month scenario of market portfolio return built with five minutes intraday data. With a complete spectrum of confidence levels from 0001 to 0.999, we analyze the VaR and the ES to see the interest of using multimodal distribution instead of unimodal distribution.

Keywords: Risks, Multimodal distributions, Value-at-Risk, Expected Shortfall, Moments method, Adapted rejection sampling, Regulation

Email addresses: `dominique.guegan@univ-paris1.fr` (Dominique Guégan), `bertrand.hassani@gmail.com` (Bertrand Hassani), `kehanleex@gmail.com` (Kehan Li)

JEL: C13 · C15 · G28 · G32

1. Introduction

In finance, assets and risks behavior can be represented using data. While in theory the distributions used to model them are usually regular and unimodal, the empirical representation are traditionally not regular and several modes may appear. In the modern portfolio theory framework the variance is used to quantify risk, and this is only valid when returns are elliptically distributed. However, if the returns follow another kind of distribution, other risk measures would be more appropriate. The variance is a symmetric measure, so it counts abnormally high returns as the same way as abnormally low returns. If investors are only interested in losses (downside risk), dispersion might be of little interest. In this paper, we only consider an asymmetric risk, and risk measurements are only useful if they reflect investors appetite, which is not necessary the case with the variance.

Consequently, features of financial data are crucial for an appropriate representation and computations of risk measures. These features - such as the support, the number of modes, the volatility clustering, the skewed property and the tail behavior - are captured by the probability distributions which characterize the data. Although the literature have been discussing features like the tail behavior of a distribution or the volatility clustering of financial data quite a lot (Jorion, 1996 [14]; Chan et al., 2007 [5]), academics and risk managers only paid little attention to the modes of a distribution when computing the risk measures. Unimodal distributions (like Gaussian or Student-t distributions) are widely used in financial institutions for Value-at-Risk (VaR) and Expected Shortfall (ES) calculations. However, when the histogram of historical data exhibits more than one modes, a unimodal distribution may provide a biased fitting of the data due to the impact of extra modes, and lead to biased risk measures. In this paper, we analyze the impact of fitting a multimodal distribution on data for risk measure computation. To fit the multimodal data various approaches will be tested, for instance the Cobb's family (Cobb et al., 1983 [6]) or the Lila family (Hassani et al., 2016 [12]).

A distribution from Cobb's family is defined in the form of a probability density, which appears as a combination of two smooth functions [1]: an exponential function and a polynomial function. The exponential function ensure that the density is positive and the polynomial function allows different shapes for the density. Cobb's family has three advantages: first, it permits the creation of multimodal distributions with various numbers of modes. The degree of the polynomial drives the number of modes of the density. The roots of the polynomial identify the locations of the modes. Besides, this family also includes classical unimodal distributions such as the Gaussian distribution, the Gamma distribution or the Beta distribution; Second, Cobb's family provides multimodal distributions with different kinds of supports: infinite support, positive support and finite support. This is useful in practice, for instance, if an histogram representing losses data is multimodal, we can fit a multimodal distribution with positive support from Cobb's family on them; Third, efficient moments estimators for the parameters of Cobb's family are provided. The estimators have a theoretical asymptotic normality property. The estimates are easy to compute because their computation relies on the empirical moments.

The main disadvantage of Cobb's family is that in general the cumulative distribution function (c.d.f) of a multimodal distribution belonging to Cobb's family does not have a closed-form expression. A closed-form expression only exists for the density. Consequently, we cannnot compute the VaR and the ES, or generate random numbers using classical inversion methods. Thus we use Monte-Carlo techniques to compute the VaR and the ES for Cobb's distribution: first, relying on a rejection sampling algorithm (Evans, 1998 [8]), a way to

[1] A smooth function is a function that has derivatives of all orders everywhere in its domain.

simulate random numbers from a multimodal probability density belonging to the Cobb's family is implemented (Gilks et al., 1992 [10]); second, these simulated random numbers are used to compute both the VaR and the ES through the empirical VaR and the empirical ES defined in section (4.1). The asymptotic results of these two empirical estimators are provided by Serfling (2001) [19] and Gao et al. (2011) [9].

Along Cobb's family, there are other ways of obtaining multimodal distributions. The first one is by using a distortion operator, which is an increasing function:

$$L : [0, 1] \to [0, 1],\ with\ L(0) = 0\ and\ L(1) = 1. \tag{1}$$

Wang (2000) [20] provides a class of distortion operators to analyze the price of risk for both insurance and financial risks. Other types of distortion operators such as the use of polynomial functions, see Guégan and Hassani (2015) [11] and Hassani and Yang (2016) [12]. An alternative way is by mixing distributions, like Gaussian mixtures (Roeder et al., 1997 [18]) and Beta mixtures (Ji et al., 2005 [13]). The mixture density is a weighted sum of various probability density components. The estimation of parameters can be done using optimization based approaches, like Expectation-maximization algorithm (Moon, 1996 [15]).

For the distortion approach, we select a well-studied unimodal distribution and "distort" it by compositing a distortion operator L and the c.d.f. of this unimodal distribution F. When we have a closed-form expression for the inverse function L^{-1} of the distortion operator, given a confidence level p, the VaR associated with the distorted distribution $L(F(x))$ is:

$$VaR_p^L = F^{-1}(L^{-1}(p)). \tag{2}$$

Also, in this case we can simulate random numbers from $L(F(x))$ by: (i) simulate random numbers from $Uniform(0, 1)$ distribution; (ii) map these numbers by $F^{-1}(L^{-1}(x))$. Then we can use these random numbers to compute the ES for $L(F(x))$.

However, sometimes we may not have a closed-form expression for L, for instance when L is a polynomial with degree greater or equal to five. Nevertheless, for the computation of VaR_p^L, we can use the Newton-Raphson method to find the root of $L(x) - p = 0$, denoted as p^L. Then $VaR_p^L = F^{-1}(p^L)$. To simulate random numbers from $L(F(x))$: (i) simulate random numbers from $Uniform(0, 1)$ distribution; (ii) for each random number r, use the Newton-Raphson method to find the root of $L(x) - r = 0$; (iii) map these roots by $F^{-1}(x)$. Then we compute the ES by simulation [2].

For the empirical analysis, after introducing both theoretical features of the distribution and algorithmic requirements, we apply the modelling strategies presented to two empirical data sets: a daily data concerning operational risk and a three month scenario of market portfolio return built with five minutes intraday data. We use the first data set to compute the VaR and the ES with one day risk horizon. The second data set allows to compute the three months VaR and ES. With a complete spectrum of confidence level p from 0.001 to 0.999, we calculate the VaR and the ES using unimodal distributions and multimodal distributions, in order to analyze the impact of multimodal distributions on risk measurement. We observe: (i) a multimodal fitting from Cobb's family provides the best goodness-of-fit; (ii) for different confidence levels, the values of the VaR and the ES associated with multimodal distributions may be larger or smaller than the values associated with unimodal distributions. In particular, the multimodal fittings from Cobb's family provide values of VaR and ES neither too high nor too low, which avoid underestimating or overestimating the risks; (iii) for operational risk data, the multimodal fitting based on the distortion approach provides the most conservative values for VaR and ES when $p = 0.999$. It means that this fitting captures the multimodal feature of the historical data,

[2]Indeed, the step (ii) may make this algorithm inefficient

and at the same time provides a fat-tail allowing to take into account the potential extreme events. Consequently, using a multimodal distribution leads to a more realistic and reasonable pricing and risk management process for financial products, and a consistent bank capital regrading their risk profile.

The remainder of the paper is structured as follows. Section 2 introduces Cobb's distribution family and illustrates it with some examples, presenting different multimodal and support behaviors. Section 3 provides a way to generate random numbers efficiently from probability densities of Cobb's family. Section 4 presents empirical study. Section 5 concludes.

2. A family of multimodal distributions

Let X be a random variable (r.v.), which may represent the daily loss of a given financial asset, portfolio or an actual incident. Let F be the cumulative distribution function of X. Traditionally, we assume X following a unimodal distribution such as a Gaussian distribution or a Student-t distribution. In this paper, we extend F from unimodal distributions to multimodal distributions. First, let us present a family of distributions (Cobb et al. (1983) [6]) and in particular its mutlimodal behavior. The general form of the Cobb's density can be expressed as follows:

$$f_k(x) = \xi(\boldsymbol{\beta}) exp \left[\int^x -\frac{g(s)}{v(s)} ds \right], \tag{3}$$

where $g(x) = \beta_0 + \beta_1 x + ... + \beta_k x^k$ and k is a positive integer. $\xi(\boldsymbol{\beta})$: $\mathcal{R}^{k+1} \to \mathcal{R}$ is the normalization function depends on $\boldsymbol{\beta} = (\beta_0, ..., \beta_k)$. This function $\xi(\boldsymbol{\beta})$ ensures that the integral of f_k over its domain equals to 1. We refer to $g(x)$ as the shape polynomial of f_k, as the maximum number of possible modes of f_k is determined by k, which is the degree of $g(x)$. Regarding the differentiation with respect to (w.r.t) x, expression (3) yields:

$$\frac{f_k'}{f_k} = -\frac{g(x)}{v(x)}, \tag{4}$$

meaning that the critical points of f_k (i.e., all x such that $f_k'(x) = 0$) are exactly the roots of $g(x)$. Indeed, each specification of $v(x)$ in (3) leads to a distinct family of distributions. We consider two types of $v(x)$ leading to two families of distributions: (i) The Gaussian family ($N_k(x)$) with $v(x) = 1$, $x \in (-\infty, \infty)$:

$$N_k(x) = \xi(\boldsymbol{\beta}) exp[\theta_1 x + \theta_2 x^2 + ... + \theta_{k+1} x^{k+1}], \tag{5}$$

where $\theta_j = -\frac{\beta_{j-1}}{j}$, $j = 1, ..., k+1$. $N_k(x)$ has finite moments of all orders if k is odd and $\theta_{k+1} < 0$. Specially, N_1 is the Gaussian density; (ii) The Gamma family ($G_k(x)$) with $v(x) = x$, $x \in (0, \infty)$:

$$G_k(x) = \xi(\boldsymbol{\beta}) x^{\alpha-1} exp[\theta_1 x + \theta_2 x^2 + ... + \theta_k x^k], \tag{6}$$

where $\alpha = 1 - \beta_0$ and $\theta_j = -\frac{\beta_j}{j}$, $j = 1, ..., k$. G_k has finite moments of all orders if $\alpha > 0$ and $\theta_k < 0$. Specially, G_1 is the Gamma density.

In practice, we use either $N_k(x)$ or $G_k(x)$ for different kinds of data. $N_k(x)$ is a benchmark. We use $G_k(x)$ when the data have a positive support, for instance operational losses.

To present the multimodal behavior of $N_k(x)$ and $G_k(x)$, we plot three graphs in Figure (1): in the first graph, we plot two densities from $N_3(x)$. The solid line has shape polynomial $g(x) = 200x(x + 0.3)(x + 0.8)$. The dash line has shape polynomial $g(x) = 1000x(x + 0.3)(x + 0.4)$; in the second graph, two densities are plotted from $G_3(x)$. The solid line has shape polynomial $g(x) = 1000(x - 0.1)(x - 0.3)(x - 0.4)$. The dash line has shape polynomial $g(x) = 100(x - 0.1)(x - 0.3)(x - 0.8)$; in the third graph, a density from $N_5(x)$ with $g(x) = 600(x + 0.2)(x + 0.5)(x + 0.55)(x + 1.2)(x + 1.25)$ has been plotted [3].

[3] Here, the parameters of $g(x)$ are chosen arbitrarily. The purpose is to show the multimodal behavior of Cobb's family, and the relationship between the locations of modes and the roots of $g(x)$.

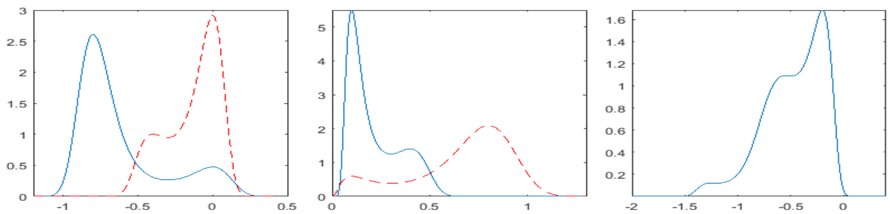

Figure 1: In the first graph, we plot two densities from $N_3(x)$. The solid line has shape polynomial $g(x) = 200x(x+0.3)(x+0.8)$. The dash line has shape polynomial $g(x) = 1000x(x+0.3)(x+0.4)$; in the second graph, we plot two densities from $G_3(x)$. The solid line has shape polynomial $g(x) = 1000(x-0.1)(x-0.3)(x-0.4)$. The dash line has shape polynomial $g(x) = 100(x-0.1)(x-0.3)(x-0.8)$; in the third graph, we plot a density from $N_5(x)$ with $g(x) = 600(x+0.2)(x+0.5)(x+0.55)(x+1.2)(x+1.25)$.

In Figure (1), all the densities in the left graph and middle graph have two modes. The density in the right graph has three modes. In the left and right graphs, the supports of densities from N_k family have both positive and negative sections. In the middle graph, the densities from G_k family only have a positive support. Moreover, in all the three graphs, we observe that the locations of the modes are exactly identified by the roots of the shape polynomials. For instance, in the left graph, the two peaks of the solid line are at 0 and -0.8 and the bottom is at -0.3. These values are just the roots of the polynomial shape $g(x) = 200x(x+0.3)(x+0.8)$.

In order to use Cobb's family in practice, we introduce a theorem in the following, which allows to fit a distribution from Cobb's family on data set. Note that the unknown parameters of Cobb's family are $\boldsymbol{\beta}$.

Theorem 1. *Let $X_1, ..., X_n$ be a sequence of losses corresponding to the previous random variable X. Then according to Cobb et al. (1983) [6], we define*

$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{M}}^{-1}\widehat{\boldsymbol{\alpha}}, \tag{7}$$

$\widehat{\boldsymbol{\beta}}$ is a consistent consistent estimator of $\boldsymbol{\beta}$ with asymptotic normality property. Where for $i, j = 1, ..., k+1$, we define $\widehat{\boldsymbol{M}} = \left[\sum_{k=1}^{n} \frac{X_k^{i+j-2}}{n}\right]_{(k+1)\times(k+1)}$, which is a positive definite random matrix with probability one. Let $\alpha_j = E\left[\left(X^{j-1}v(X)\right)'\right]$, then we define a vector $\widehat{\boldsymbol{\alpha}} = [\widehat{\alpha}_j]_{(k+1)\times 1}$ where $\widehat{\alpha}_j$ is the associated sample moments of α_j. $\widehat{\alpha}_j$ depends on the type of $v(x)$, for example, for $N_k(x)$ with $v(x) = 1$, $\widehat{\alpha}_j = \sum_{k=1}^{n}(j-1)X_k^{j-2}$.

The proof of Theorem (1) is provided in Cobb et al. (1983) [6]. Following this theorem, we can use moments estimator to estimate $\boldsymbol{\beta}$ in practice.

3. An algorithm for random sample generation from Cobb's family

In this section, we develop an algorithm to generate random samples from $f_k(x)$ given by equation (3), based on the rejection sampling technique introduced by Evans (1998) [8]. Efficient algorithms for generating random samples from probability densities is a necessary part of many applications related to integral approximation, such as approximating the mean of a r.v. by Monte-Carlo simulation according to the law of large numbers. In the next section, we use these simulated random samples to compute the VaR and the ES for multimodal distributions belonging to Cobb's family. As an example, we choose the distribution $N_k(x)$ in equation (5) as our target density and explain how to generate random samples from it using following algorithm:

3.1. The algorithm

There are three steps in the algorithm.

- Step 1: it is equivalent to perform the rejection sampling by $N_k(x)/\xi(\boldsymbol{\beta})$ instead of $N_k(x)$ itself (Gilks (1992) [10]). We use the logarithm to transform $N_k(x)/\xi(\boldsymbol{\beta})$ into $N_k^*(x) = log(N_k(x)/\xi(\boldsymbol{\beta})) = \theta_1 x + \theta_2 x^2 + ... + \theta_{k+1}x^{k+1}$. Initially, we need to introduce a set of points $a \leq x_1 < ... < x_m \leq b$. All the critical and inflection points of $N_k^*(x)$ should be included

[4]. A Newton-Raphson approach is implemented to numerically solve the equations $N_k^{*\prime}(x) = 0$ and $N_k^{*\prime\prime}(x) = 0$ to find these points. Points such that $t_i(x) = 0$ should be avoided.

- Step 2: an upper envelop function for $N_k^*(x)$ is established. Let $t_i(x)$ be the tangent line to $N_k^*(x)$ at x_i, i.e.

$$t_i(x) = N_k^*(x_i) + N_k^{*\prime}(x_i)(x - x_i), \tag{8}$$

we define $x_i < z_i < x_{i+1}$, $1 \leq i \leq m-1$ as the point of intersection for two neighbouring tangent lines (i.e. $t_i(z_i) = t_{i+1}(z_i)$) [5] and put $z_0 = a$, $z_m = b$. Furthermore, let $c_i(x)$ be the secant line to $N_k^*(x)$ from $(z_{i-1}, N_k^*(z_{i-1}))$ to $(z_i, N_k^*(z_i))$, i.e.

$$c_i(x) = N_k^*(z_i) + (N_k^*(z_i) - N_k^*(z_{i-1}))(x - z_i)/(z_i - z_{i-1}), \tag{9}$$

then the upper envelop function of $N_k^*(x)$ is defined by:

$$u(x) = \begin{cases} t_i(x), & if \quad z_{i-1} \leq x \leq z_i \quad and \quad \frac{z_{i-1}+z_i}{2} \leq 0, \\ c_i(x), & if \quad z_{i-1} \leq x \leq z_i \quad and \quad \frac{z_{i-1}+z_i}{2} \geq 0. \end{cases} \tag{10}$$

By definition, we have $N_k^*(x) \leq u(x)$ for $\forall x \in (a, b)$. Since the exponential function is monotonic, we have $N_k(x)/\xi(\boldsymbol{\beta}) \leq e^{u(x)}$, where $e^{u(x)}$ is a piecewise exponential function. We define $h(x) = e^{u(x)}/\int_a^b e^{u(x)}dx = \sum_{i=1}^m p_i h_i(x)$, where $d_i = \int_{z_{i-1}}^{z_i} e^{u(x)}dx$, $p_i = d_i/\sum_{i=1}^m d_i$ and $h_i(x) = e^{u(x)}/d_i$ on $[z_{i-1}, z_i]$ and is equal to 0 elsewhere. We use the composition method of Devroye (1986) [7] to obtain random samples from $h(x)$: we

[4] For a function f, x_1 is a critical point for f if $f'(x_1) = 0$. x_2 is an inflection point for f if $f''(x_2) = 0$ and $f'''(x_2) \neq 0$. In our case, the critical points are also the real roots of the shape polynomial.

[5] z_i exists because $N_k^*(x)$ is either concave or convex in (x_i, x_{i+1}).

generate an integer z in $\{1, ..., m\}$ with discrete probability $P(Z = z) = p_z$; we use the inversion method to generate a random sample x with density h_z [6]. Then x is a sample from $h(x)$.

- Step 3: we perform the rejection sampling algorithm as following: (i) generating x from $h(x)$; (ii) generating v from $U(0, 1)$ (uniform density on $(0, 1)$); (iii) if $N_k^*(x) \geq v * u(x)$, then return x else go to (i).

By definition of rejection sampling, the accepted x follows $N_k(x)$. When the acceptance rate α is low, where $\alpha = P(N_k^*(X) \geq V * u(X)) = \frac{\int_a^b N_k^*(x)dx}{\int_a^b u(x)dx}$, we can also work in an adaptive way, i.e. we come back to step 1 and add the rejected point x to the initial set $\{x_1, ..., x_m\}$ each time. Indeed the sequence of $u(x)$ converge to $N_k^*(x)$ when this procedure is iterated. As $u(X)$ becomes closer to $N_k^*(x)$, α grows (Martino, 2011 [16]). To stop the adaptive procedure, we need to choose a value of α a prior.

3.2. An implementation of the algorithm

We illustrate how the algorithm in section (3.1) performs considering the distribution $N_3(x)$ with $g(x) = 1000x(x + 0.3)(x + 0.4)$. The critical points are 0, -0.3 and -0.4. The inflection points are -0.1131 and -0.3535. Noticing that $t_i(x) = 0$ when $x_i = 0$, we choose the initial partition: $\{-1, -0.4, -0.3535, -0.3, -0.1131, 0.005, 1\}$. In order to increase the acceptance rate of the rejection sampling algorithm, we add points to the initial partition to decrease the space between the upper envelop and the transformed target density. In Figure (2): the dash lines are the upper envelops and the solid line is the transformed target density. The plot on the left corresponds to the initial partition. For the graph in the middle, the lattice is enlarged as we add 0.4

[6] Notice that h_z is an exponential function having closed-form for integral and inverse mapping.

to the initial partition. The plot on the left is obtained with the partition $\{-1, -0.5, -0.4, -0.3535, -0.3, -0.1131, -0.05, 0.005, 0.05, 0.1, 0.3, 0.4, 1\}$. We observe that the upper envelop converges to the transformed target density as we add points to the partition.

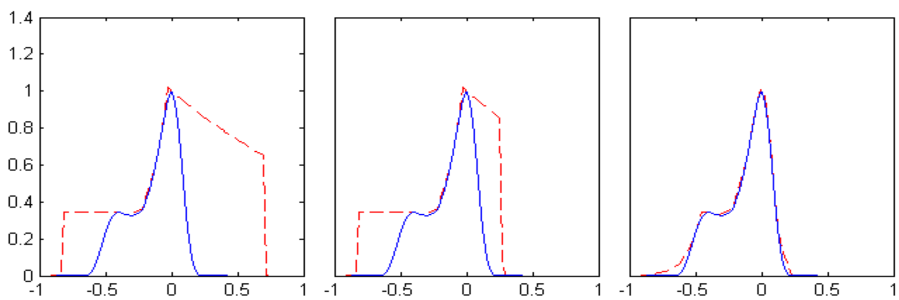

Figure 2: The dash lines are the upper envelop and the solid lines are the transformed target density. The left graph corresponds to the initial partition. For the graph in the middle, we add 0.4 to the initial partition. The right graph is with the partition $\{-1, -0.5, -0.4, -0.3535, -0.3, -0.1131, -0.05, 0.005, 0.05, 0.1, 0.3, 0.4, 1\}$.

Since we find an upper envelop which is close to the transformed target density, the rejection sampling algorithm can be implemented efficiently. We simulate $n = 1000, 10000, 100000$ random samples from $N_3(x)$ with $g(x) = 1000x(x+0.3)(x+0.4)$ using rejection sampling, and show the results in Figure (3): the solid line is our target density; the histogram on the left of Figure (3) corresponds to the sample size $n = 1000$; the histogram in the middle corresponds to $n = 10000$ and the histogram on the right has been obtained considering $n = 100000$. We observe that samples obtained from our algorithm reach the target density when n increases.

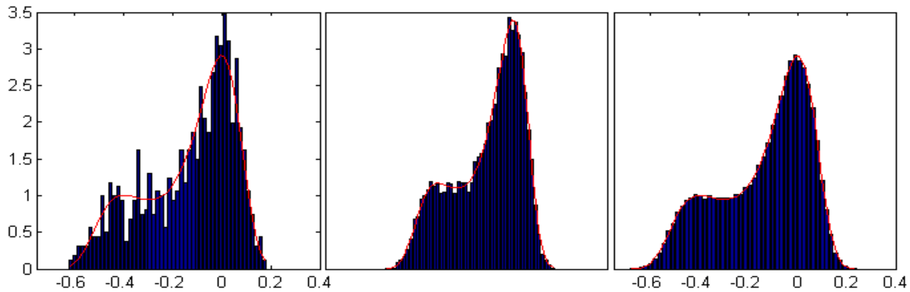

Figure 3: The histogram on the left corresponds to the samples from the rejection sampling with sample size $n = 1000$. The histogram in the middle corresponds to $n = 10000$ and the histogram on the right has been been obtained considering $n = 100000$. We observe that samples obtained with our algorithm indeed reach the target density.

4. Empirical study: the impact of multimodal distributions on risk measurement

Among others, quantile-based risk measures are the most prominent instruments used by banks and supervisors, for either day-to-day internal risk management purposes or regulatory capital requirements. Consequently, it is critical to accurately compute these risk measures for data with different features, in order to promote the stability and efficiency of the financial system and contribute to the real economy. The VaR summarises the worst potential loss over a target risk horizon within a given confidence level. The ES is defined as the expected loss beyond the VaR with a given risk horizon and confidence level.

4.1. Tools

Let F^{-1} be the left continuous inverse of X's c.d.f F, i.e. $F^{-1}(x) = min\{u : F(u) \geq x\}$. For a given confidence level $0 \leq p \leq 1$, we define the Value-at-Risk VaR_p as the p-quantile :

$$VaR_p = F^{-1}(p). \tag{11}$$

For the same confidence level p, the Expected Shortfall ES_p is equal to:

$$ES_p = \frac{1}{1-p}\int_p^1 F^{-1}(u)du = \frac{1}{1-p}\int_p^1 VaR_u du. \tag{12}$$

Although in the new standards of minimum capital requirements for market risk (BCBS, 2016 [3]), there is a proposal to shift from VaR to an ES measure of risk under stress, we consider both the VaR and the ES in this paper. Because the VaR is still one of the most popular risk measures among financial institutions (Pérignon, 2010 [17]).

In order to compute the VaR and the ES of a multimodal distribution, we introduce the empirical VaR and the empirical ES in the following: we rank $X_1, ..., X_n$ and obtain $X_{(1)} \leq \cdots \leq X_{(n)}$ and define the empirical VaR as:

$$eVaR_p = X_{(m)}, \tag{13}$$

where $m = np$ if np is an integer and $m = [np] + 1$ otherwise. $[x]$ denotes the largest integer less than or equal to x.[7]. Furthermore, $eVaR_p$ is a consistent estimator of VaR_p (Serfling, 2009 [19]).

Based on the $eVaR$, we define the empirical Expected Shortfall as

$$eES_p = \frac{m-np}{n(1-p)}X_{(m)} + \frac{1}{n(1-p)}\sum_{i=m+1}^{n} X_{(i)}. \tag{14}$$

It is a consistent estimator of the ES_p (Acerbi 2002 [1] and Gao 2011 [9]). Brazauskas (2008) [4] obtained strong consistency and asymptotic normality of the eES_p. In the following, we use these empirical estimators and the algorithm in section 3.1 to compute the VaR and the ES for empirical analysis, using an operational risk data set and a market risk data set.

[7] $X_{(m)}$ is also called the mth order statistic, which is a fundamental tool in nonparametric statistics.

Besides Cobb's family, we also consider the distortion family as alternative multimodal distributions. Following Hassani et al. (2016) [12], we consider the general Double-Odd-Polynomial Seat function as a distortion operator:

$$L(x) = \begin{cases} b - b(1 - \frac{x}{a})^{2n+1}, & x \le a, \\ b + (1-b)(\frac{x-a}{1-a})^{2n+1}, & x > a. \end{cases} \tag{15}$$

Where $0 < a < 1$, $0 \le b \le 1$ and integer $n \le 1$. We derive the inverse function of $L(x)$:

$$L^{-1}(x) = \begin{cases} a - a \sqrt[2n+1]{1 - \frac{x}{b}}, & 0 \le x \le b, \\ a + (1-a) \sqrt[2n+1]{\frac{x-b}{1-b}}, & b < x \le 1. \end{cases} \tag{16}$$

Hassani et al. (2016) [12] show that $L(x)$ can transform a unimodal distribution to a multimodal distribution. We use this approach as an alternative multimodal distribution family to compare with Cobb's family in the following application.

4.2. Data description

We consider two data sets:

1. For the first data set Data1: (i) we download the valuation of CSI 300 index every 5 minutes from Bloomberg. The CSI 300 index is a free-float weighted index that consists 300 A-share stocks listed on the Shanghai or Shenzhen Stock Exchanges, China; (ii) the log-return of the downloaded data is computed; (iii) a rolling window with length 3000 is used to obtain a three month log-return scenario of this index. The Data1 has 65327 data.

2. The second data set Data2 is provided by a European bank representing "External Fraud Losses" risks since 2013 on a daily basis. "External Fraud" risk is a sub-category of operational risk.

The first four empirical moments of Data1 and Data2 are provided along the number of observations in Table (1). In Data1, the minimum log-return is -0.2369 and the maximum log-return is 0.1832. In Data2, the minimum loss is 2032.7 and the maximum loss is 40438. From Table (1), we observe that both Data1 and Data2 are right skewed. The empirical kurtosis of Data1 is 4.2392 and the empirical kurtosis of Data2 is 4.0112. These values are slightly larger than 3, which means that we do not expect a tremendous fat-tailed behavior of the data. However, we show in the following that these two data sets contain information in the tails.

		Empirical moments			
	points	mean	variance	skewness	kurtosis
Data1	65327	0.0035	0.0038	0.1746	4.2392
Data2	1000	1.1129e+04	3.7242e+07	1.0768	4.0112

Table 1: In this table, we provide the first four empirical moments of Data1 and Data2 and the number of observations.

4.3. Application

In order to compare the performances of the unimodal distribution and multimodal distribution, we perform two parts empirical analysis: the first part concerning the market data Data1 and the second part concerning the operational risk data Data2.

4.3.1. Market data

In practice, the market data may present multimodal features. For instance, traders may create portfolios containing assets with different properties to pursue the diversification bonus. Consequently, these portfolios contain different sources of information from the market. Then the histogram of the Profit & Loss data for these portfolios may have several modes.

4.3.1.1 Fittings and adequacy

For Data1, a Student-t distribution and a Normal-inverse Gaussian distribution (NIG; Barndorff-Nielsen, 1978 [2]) are fitted using maximum likelihood approach; a Student-t distribution distorted by $L(x)$ in equation (15) with $n = 1$ is fitted using quantile distance minimisation method; a N_5 distribution is fitted using moments method as defined in Theorem (1). The estimates are provided in Appendix A.

To analyze the goodness-of-fit, two distances are computed: the first one is the two sample Kolmogorov–Smirnov distance (DKS), which is the maximum distance between the empirical c.d.f of two data sets; the second one, which is the sum of distances at different quantiles between the empirical c.d.f of two data sets (S-DKS). The results are provided in Table (2).

	Student-t	NIG	N_5
DKS	0.012	0.009	0.0049
S-DKS	3.9453	6.9458	2.7312

Table 2: In this table, we provide the values of the DKS and the S-DKS for different fittings of Data1.

From Table (2), the N_5 fitting from Cobb's family provides the best goodness-of-fit. Because it provides the smallest DKS and S-DKS which are equal to 0.0049 and 2.7312.

We provide a figure to illustrate the fittings on Data1: in Figure (4), we plot the histogram of Data1. The dash line represents the N_5 fitting; the dot-dash line represents the Student-t fitting; the solid line represents the NIG fitting.

In Figure (4): (i) the highest peak is located between -0.05 and 0.05. All the three fittings capture this part of information; (ii) there is a second peak in

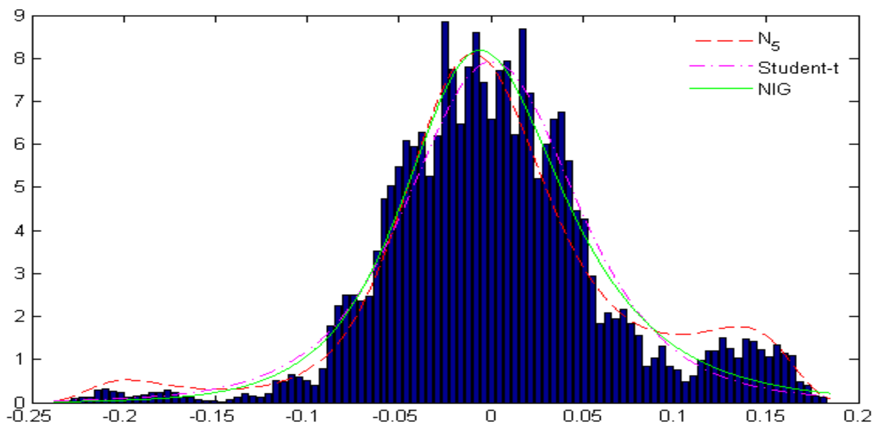

Figure 4: We plot the histogram of Data1. The dash line represents the N_5 fitting; the dot-dash line represents the Student-t fitting; the solid line represents the NIG fitting.

the right tail, between 0.11 and 0.17. This part of information represents the potential gain of this portfolio. It may be one reason for traders to hold this portfolio. However, only the N_5 fitting, which has the second mode between 0.11 and 0.17, takes this piece of information into account. On the contrary, the other two unimodal fittings ignore this piece of information because they quickly decrease between 0.11 and 0.17; (iii) more importantly, we also observe a third peak in the left tail between -0.16 and -0.24. The information in this peak represents the potential loss of the portfolio. According to the assumption of no arbitrage, the peak in the left tail can be viewed as a reasonable compensation of the peak in the right tail. Therefore, it is very important for risk managers not to miss that part when they measure the risks of this portfolio. It is also important for traders to decide if they should hold a position in this portfolio or not. Indeed, from the figure, we can see that only the N_5 fitting, which exhibits the third mode between -0.16 and -0.24, captures the information in the third peak of the histogram. The other two unimodal fittings quickly decrease in this area.

4.3.1.2 Spectrum for market data

In the following, with a complete spectrum of confidence levels p ranging from 0.001 to 0.999, we compute the VaR and the ES with three month horizon for Data1, using the fittings in Appendix A. Given $0 < p < 1$, 10000 random numbers are simulated from each fitted distribution and the $eVaR_p$ and the eES_p are computed. This approach is repeated 100 times. Then 100 values of $eVaR_p$ and 100 values of eES_p are obtained. We take the average of $eVaR_p$ and eES_p as the values of the VaR_p and the ES_p.

We illustrate the results for the values of the VaR_p and the ES_p graphically providing Figure (5): the results for the VaR are presented in the top graph and the results for the ES are presented in the bottom graph. The solid line represents the VaR and the ES values computed from the Student-t fitting; the dot-solid line represents the VaR and the ES values computed from the NIG fitting; the dash line represents the VaR and the ES values computed from the N_5 fitting.

From the top graph of Figure (5): (i) for $0.33 \leq p \leq 0.48$ and $0.005 \leq p \leq 0.05$, the N_5 fitting provides the smallest values of the VaR. Because it captures the piece of information in the peak around -0.2. However, the unimodal fittings ignore this information and put more weights at -0.1. Furthermore, for $p = 0.001$, the VaR value of the Student-t fitting is equal to -0.3045, which is smaller than the VaR value of the N_5 fitting -0.2265. Because the Student-t fitting put weights below the minimum of Data2 -0.2369; (ii) for $0.65 \leq p \leq 0.975$, the N_5 fitting provides the largest VaR values. It is because the N_5 fitting takes into account the peak associating with the profit of the portfolio. Nevertheless, for $0.99 \leq p \leq 0.999$, the N_5 fitting gives the smallest VaR values. Instead, the Student-t fitting provides the largest VaR values. For instance, $VaR_{0.999}$ is equal to 0.3052 for the Student-t fitting, which is larger than the maximum of Data2 0.1832. But $VaR_{0.999}$ is equal to 0.1781 for the N_5 fitting. The reason

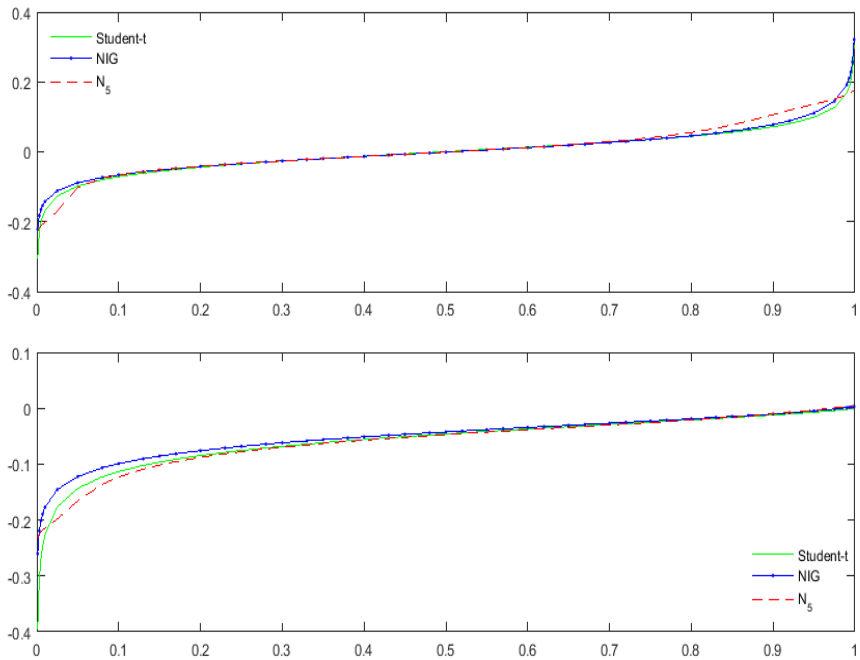

Figure 5: For Data1, we illustrate the results for the values of the VaR_p and the ES_p graphically in this figure. The results for the VaR are presented in the top graph and the results for the ES are presented in the bottom graph. The solid line represents the VaR and the ES values computed from the Student-t fitting; the dot-solid line represents the VaR and the ES values computed from the NIG fitting; the dash line represents the VaR and the ES values computed from the N_5 fitting.

is that the Student-t fitting shifts the piece of information in the peak around 0.15 to the area beyond 0.1832.

From the bottom graph in Figure (5): (i) $0.025 \leq p \leq 0.8$, the N_5 fitting provides the smallest values of the ES. It is conservative because it captures the information in the peak around -0.2. The other two unimodal fittings ignores this information; (ii) for $p = 0.001$, the value of the ES for the Student-t fitting is equal to -0.395, which is nearly twice of the ES value for the N_5 fitting.

Indeed, the Student-t fitting shifts the information in the peak around -0.2 to the area below -0.2369; (iii) for $0.92 \leq p \leq 0.999$, the N_5 fitting gives the largest values of the ES, because it takes into account the peak around 0.15.

4.3.2. Operational risk data

As the operational risk data may contain different kinds of operational losses coming from different activities of banks, they may exhibit multimodal behavior.

4.3.2.1 Fittings and adequacy

For Data2, a Log-normal distribution, a NIG distribution, a lognormal distribution distorted by $L(x)$ in equation (15) with $n = 1$ and a G_7 distribution are fitted. Because Data2 has a positive support. We use the same estimation techniques for Data2 as we use for Data1. The estimates are provided in Appendix A.

To analyze the goodness-of-fit, we compute the DKS and S-DKS and provide the results in Table (3).

	Lognormal	NIG	D-Lognormal	G_7
DKS	0.0424	0.0615	0.2948	0.0154
S-DKS	10.3246	24.6571	129.1102	4.9859

Table 3: In this table, we provide the values of the DKS and the S-DKS for different fittings of Data2.

From Table (3), for Data2, the smallest DKS is 0.0154 obtained with the G_7 fitting. The second smallest DKS is obtained with the Lognormal distribution and is equal to 0.0424. It is more than twice as large as the DKS value obtained using the G_7 fitting. The results for the S-DKS are consistent with the results of DKS: the G_7 fitting provides the smallest S-DKS value 4.9859. The largest DKS and S-DKS are obtained with the D-Lognormal fitting and are equal to 0.2948 and 129.1102. They are nearly five times as large as the values obtained using the NIG fitting, which provide the second largest DKS and S-DKS. Therefore, on Data1, the G_7 fitting provides the best goodness-of-fit and the D-Lognormal fitting provides the worst goodness-of-fit in the sense of DKS and S-DKS.

We provide a figure to illustrate the fitting results for Data2: in Figure (6), we plot the histogram of Data2. The dash line represents the G_7 fitting; the dot-dash line represents the Lognormal fitting; the solid line represents the NIG fitting; the dot line represents the distorted Lognormal fitting

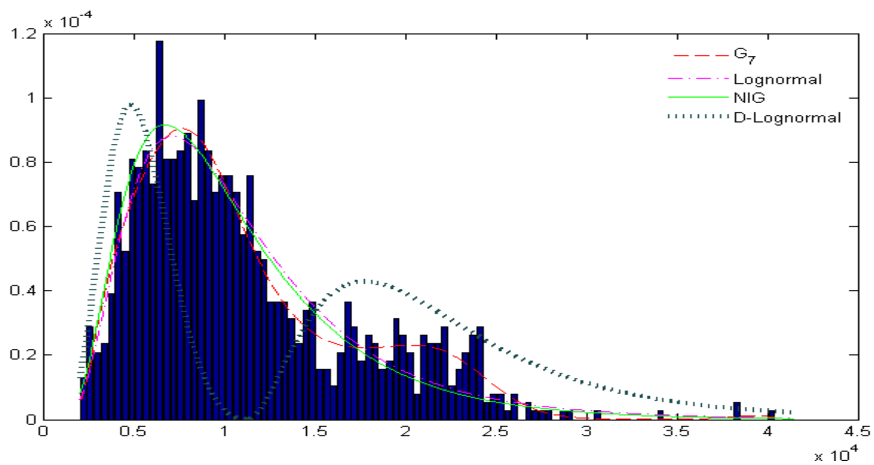

Figure 6: We plot the histogram of Data2. The dash line represents the G_7 fitting; the dot-dash line represents the Lognormal fitting; the solid line represents the NIG fitting; the dot line represents the distorted Lognormal fitting.

In Figure (6): (i) the highest peak is between 5×10^3 and 1.4×10^4. The G_7, Lognormal and NIG fittings capture this piece of information. But the D-Lognormal fitting shifts part of information between 8×10^3 and 1.2×10^4 to the tail; (ii) we also observe that the second peak in the histogram is between 1.6×10^4 and 2.4×10^4. And here, both the G_7 and D-Lognormal fittings, which have another mode in this area, capture the information contained in the second peak. Furthermore, the D-Lognormal fitting puts more weights in the second peak than the G_7 fitting. The other two unimodal fittings decrease fast in the area between 1.6×10^4 and 2.4×10^4, meaning that the information in the second peak are ignored by these two fittings. (iii) the second peak provides important information in the tail, even though the empirical kurtosis of Data2 is just 4.0112. It means that the empirical kurtosis may not be a sensitive measure for the extra peaks in the tail of a distribution. However, the information in these peaks may be important for investors and risk managers to assess potential risks, especially "clustering of tail events" correctly.

4.3.2.2 Spectrum for operational risk data

In the following, with a complete spectrum of confidence levels p ranging from 0.001 to 0.999, we compute the VaR and the ES with daily horizon for Data2, using the fittings in Appendix A. The approach in section (4.3.1.2) is used for the computation.

We illustrate the results for the values of the VaR_p and the ES_p graphically. On Figure (7): the results for the VaR are presented in the top graph and the results for the ES are presented in the bottom graph. The solid line represents the VaR and the ES values computed from the Lognormal fitting; the dot-solid line represents the VaR and the ES values computed from the NIG fitting; the dash line represents the VaR and the ES values computed from the D-Lognormal fitting; the dot line represents the VaR and the ES values computed from the G_7 fitting.

From the top graph in Figure (7): (i) for $0.001 \leq p \leq 0.38$, the D-Lognormal fitting provides the smallest values of the VaR. However, for $0.43 \leq p \leq 0.999$, the D-Lognormal fitting provides the largest values of the VaR. Indeed, between $p = 0.4$ and $p = 0.43$, there is a jump for the VaR values of D-Lognormal fitting:

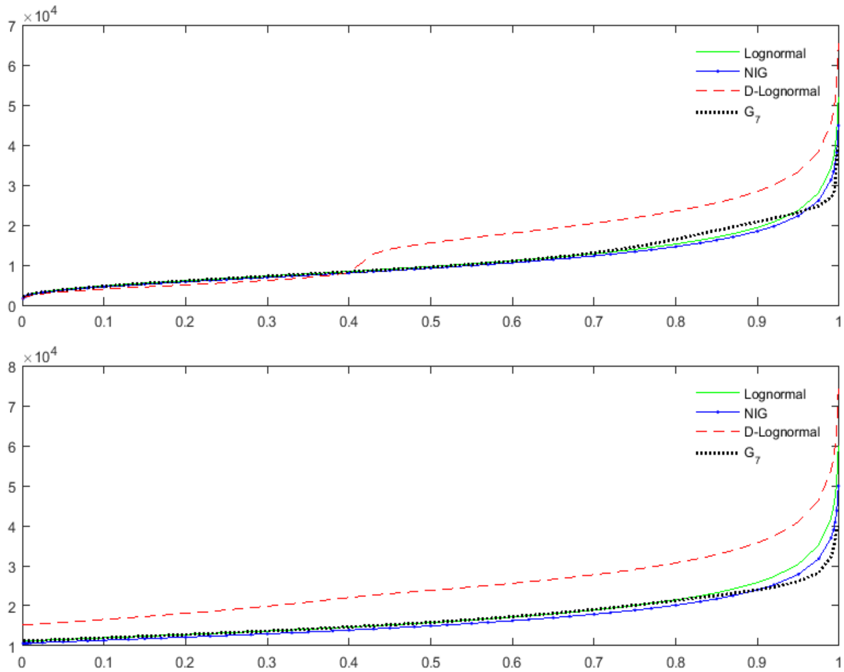

Figure 7: For Data2, we illustrate the results for the values of the VaR_p and the ES_p graphically in this figure. The results for the VaR are presented in the top graph and the results for the ES are presented in the bottom graph. The solid line represents the VaR and the ES values computed from the Lognormal fitting; the dot-solid line represents the VaR and the ES values computed from the NIG fitting; the dash line represents the VaR and the ES values computed from the D-Lognormal fitting; the dot line represents the VaR and the ES values computed from the G_7 fitting.

$VaR_{0.4}$ is 8126 and $VaR_{0.43}$ is 12789. It's because the D-Lognormal fitting shifts most of the information between 8126 and 12789 to the area beyond 12789; (ii) for $0.65 \leq p \leq 0.92$, the G_7 fitting gives the second largest values of the VaR. Nevertheless, for $0.99 \leq p \leq 0.999$, the G_7 fitting provides the smallest values of the VaR. Between $p = 0.95$ and $p = 0.99$, we observe jumps for the VaR values of Lognormal fitting from 23650 to 34324, and the VaR values of NIG fitting from 22256 to 31314.

From the bottom graph in Figure (7) associated with the ES results: (i) for all $0 < p < 1$, the D-Lognormal fitting provides the largest Values of the ES. Thus it associates with the most conservative values of ES. The reason of the conservative is that the distortion operator shifts the weights in the center to the far tail. Notice that $ES_{0.999}$ is equal to 74499 for the D-Lognormal fitting, which is nearly twice as large as the maximum of Data1 40438. That means the ES provided by the D-Lognormal fitting may be over conservative; (ii) for $0.05 \leq p \leq 0.77$, the G_7 fitting provides the second largest values of the ES. But for $0.92 \leq p \leq 0.999$, the G_7 fitting provides the smallest values of the ES; (iii) for the unimodal fitting, the Lognormal fitting ignores the information in the second peak by decreasing fast between 16000 and 40000. It leads to the values of ES for Lognormal fitting jump from 51298 to 60168 between $p = 0.997$ and $p = 0.999$.

5. Conclusion

In this paper, we discuss the necessity of considering multimodal distribution as an alternative for Value-at-Risk and Expected Shortfall computation. We compare two types of multimodal distributions - Cobb's family and distorted family - with several unimodal distributions. An adapted rejection sampling approach is proposed for generating random numbers from Cobb's family, in order to compute the VaR and the ES by Monte-Carlo simulation. For the distortion family, we suggest using the inversion method for risk measure computation. Given the full spectrum of confidence levels from 0.001 to 0.999, by computing the VaR and the ES for an operational risk data set and a market data set with daily and three month risk horizons, our empirical study shows that: first, the Cobb's family provides the best fitting in the sense of Kolmogorov-Smirnov distance. However, the distortion family provides the worst fitting; second, the curves of the VaR and the ES corresponding to multimodal fittings are sometimes above and sometimes below the others; third, the values of risk measures computed from the Cobb's family neither go too high or too low. Furthermore, they are bounded by the maximum and minimum of the historical data. That means if the risk measure is used for regulatory or economic capital purposes, then using a multimodal distribution may help generating a charge consistent with the risk profile of a bank; finally, the multimodal fitting from the distorted approach can capture the multimodal behavior of the historical data, and at the same time consider the information in the far tail. It allows to provide conservative risk measures. We understand that the use of parametric multimodal distribution is still at its infancy, however, we believe that building a set rules relying on this tool might help improving risk management.

6. Acknowledgments

This work was achieved through the Laboratory of Excellence on Financial Regulation (Labex ReFi) supported by PRES heSam under the reference ANR10LABX0095. It benefited from a French government support managed by the National Research Agency (ANR) within the project Investissements d'Avenir Paris Nouveaux Mondes (investments for the future Paris New Worlds) under the reference ANR11IDEX000602.

References

[1] Acerbi, C., Tasche, D. (2002) On the coherence of expected shortfall. Journal of Banking & Finance, 26(7), 1487–1503.

[2] Barndorff-Nielsen, O. (1978) Hyperbolic distributions and distributions on hyperbolae. Scandinavian Journal of Statistics, 5(3), 151-157.

[3] Basel Committee on Banking Supervision (2016) Minimum capital requirements for market risk. Bank for International Settlements, Basel, Switzerland.

[4] Brazauskas, V., Jones, B.L., Puri, M.L., Zitikis, R. (2008) Estimating conditional tail expectation with actuarial applications in view. Journal of Statistical Planning and Inference, 138(11), 3590–3604.

[5] Chan, N.H., Deng, S.J., Peng, L., Xia, Z. (2007) Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. Journal of Econometrics, 137(2), 556-576.

[6] Cobb, L., Koppstein, P., Chen, N.H. (1983) Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family. Journal of the American Statistical Association, 78(381), 124-130.

[7] Devroye, L. (1986) Non-Uniform Random Variate Generation. Springer-Verlag New York.

[8] Evans, M., Swartz, T. (1998) Random Variable Generation Using Concavity Properties of Transformed Densities. Journal of Computational and Graphical Statistics, 7(4), 514-528.

[9] Gao, F., Wang, S. (2011) Asymptotic behavior of the empirical conditional value-at-risk. Insurance: Mathematics and Economics, 49(3), 345–352.

[10] Gilks, W.R., Wild, P. (1992) Adaptive Rejection Sampling for Gibbs Sampling. Journal of the Royal Statistical Society. Series C (Applied Statistics), 41(2), 337-348.

[11] Guégan, D., Hassani, B. (2015) Distortion Risk Measure or the Transformation of Unimodal Distributions into Multimodal Functions. In: Bensoussan, A., Guégan, D., Tapiero, C. (ed) Future Perspective in Risk Models and Finance, 1st edn. Springer International Publishing, New York, pp. 71-88.

[12] Hassani, B., Yang, W. (2016) The Lila distribution and its applications in risk modelling. Working paper, 2016.68, Documents de travail du Centre d'Economie de la Sorbonne.

[13] Ji, Y., Wu, C., Liu, P., Wang, J., Coombes, K.R. (2005) Applications of beta-mixture models in bioinformatics. Bioinformatics, 21(9), 2118–2122.

[14] Jorion, P. (1996) Risk2: Measuring the risk in value at risk. Financial Analysts Journal, 52(6), 47-56.

[15] Martino, L., Míguez, J. (2011) A generalization of the adaptive rejection sampling algorithm. Statistics and Computing, 21(4), 633-647.

[16] Moon, T.K. (1996) The expectation-maximization algorithm. IEEE Signal Processing Magazine, 13(6), 47-60.

[17] Pérignon, C., Smith, D.R. (2010) The level and quality of Value-at-Risk disclosure by commercial banks. Journal of Banking & Finance, 34(2), 362-377.

[18] Roeder, K., Wasserman, L. (1997) Practical Bayesian Density Estimation Using Mixtures of Normals. Journal of the American Statistical Association, 92(439), 894-902.

[19] Serfling, C.J. (2009) Approximation Theorems of Mathematical Statistics. John Wiley & Sons, Hoboken, New Jersey.

[20] Wang, S.S. (2000) A Class of Distortion Operators for Pricing Financial and Insurance Risks. The Journal of Risk and Insurance, 67(1), 15-36.

Appendix A. Estimates of the distributions for Data1 and Data2

In order to compare the performances of the unimodal distribution and multimodal distribution: (i) For Data1, a Student-t distribution and a Normal-inverse Gaussian distribution (NIG; Barndorff-Nielsen, 1978 [2]) are fitted using maximum likelihood approach; a Student-t distribution distorted by $L(x)$ in equation (15) with $n = 1$ is fitted using quantile distance minimisation method; a N_5 distribution is fitted to Data1 using moments method as defined in Theorem (1); (ii) For Data2, a Log-normal distribution, a NIG distribution, a lognormal distribution distorted by $L(x)$ in equation (15) with $n = 1$ and a G_7 distribution are fitted. Because Data2 has a positive support. We use the same estimation techniques for Data2 as we use for Data1. The results are provided in Table (A.4).

Data1								
	location	scale	shape					
Student-t	0.0004	0.0477	4.5204					
	tail	skewness	location	scale				
NIG	18.3885	4.32	-0.0126	0.066				
	n	a	b	df	location	scale		
D-Student-t	2	N/A	N/A	N/A	N/A	N/A		
	β_0	β_1	β_2	β_3	β_4	β_5		
N_5	6.8301	668.1	-4.396e+3	-6.658e+4	2.0518e+5	1.7579e+6		
Data2								
	location	scale						
Lognormal	9.1721	0.5466						
	tail	skewness	location	scale				
NIG	0.0116	0.0115	-204.0365	1756				
	n	a	b	location	scale			
D-Lognormal	1	0.6143	0.4239	9.1714	0.5461			
	n	a	b	location	scale			
D-NIG	1	N/A	N/A	N/A	N/A			
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
G_7	-28.39	0.019	-5.156e-6	7.2e-10	-5.365e-14	2.141e-18	-4.286e-24	3.365e-28

Table A.4: We provide the fitted parameters for Data1 and Data2.