Technical reserves and solvency capital of insurance company: how to use the Value-at-Risk?

Pierre-E. THÉROND[α] **Frédéric PLANCHET**[*]

ISFA – Laboratoire SAF
Université Claude Bernard Lyon 1 [β]
WINTER & Associés [γ]

ABSTRACT

The advent of the future European prudential framework (Solvency II) and, to a lesser extent, of the phase II of the IFRS dedicated to the insurance contracts, will systematize the use of the Value-at-Risk (VaR) risk measure in insurance. Especially used for financial purposes, the measure of an insurance risk by a VaR technique requires a specific adaptation. Schematically we may discern two different contexts, which impose distinct approaches:

- the measurement of the risk related to the sinistrality by the use of a VaR in the heart of the probability distribution : the technical provision will have to be enough to pay the claims with a 75 % probability ;
- the measurement of risk related to the ruin of the company by the mean of a very high order VaR : the solvency capital must to be calibrated to control the ruin if the insurer with a probability higher than 99.5 %.

In the first situation, the two standard approaches (historical VaR or modeling the sinistrality) can be based on a statistical material of relative but sufficient size to estimate a VaR in the heart of the probability distribution. In the second case, we are confronted to the absence of observations. Also we have to model the basic variables which influence the solvency of the company in order to be able to simulate the ruin of the company and finally to estimate the high order VaR. This last stage will require the use of Extreme Value Theory.

In this paper, we present the contexts of VaR computation in insurance, the related probabilistic results and the limits of these kinds of criterion for insurance purposes.

[α] Pierre Thérond is a Ph.D. Student and is teaching at the actuarial school ISFA (Université Claude Bernard Lyon 1 – France). He works as an actuary for the consulting company WINTER & Associés. Contact : ptherond@winter-associes.fr

[*] Frédéric Planchet is an associate Professor of finance and insurance at the actuarial school ISFA. He is an Associate Actuary of the consulting company WINTER & Associés. Contact : fplanchet@winter-associes.fr

[β] Institut de Science Financière et d'Assurances (ISFA) - 50 avenue Tony Garnier - 69366 Lyon Cedex 07 – France.

[γ] WINTER & Associés – 43-47 avenue de la Grande Armée - 75116 Paris et 18 avenue Félix Faure - 69007 Lyon – France.

A few years ago, the banking sector experienced the advent of the so-called Basel II. Currently, the European insurance sector is getting ready to endow itself with a prudential framework resulting from the Solvency II project. Also outlined on a three-pillar construction, the Solvency system will get the insurance firms to measure and manage their risks the better. A well-managed risk might engender a lower capital requirement. The first pillar, focusing on the quantitative requirements, amply refers to the well known financial risk measurements that are the Value-at-Risk (VaR). The work in progress on Solvency II allows the insurers to own enough technical reserves[1] to meet their commitment to pay the benefits with a probability of 75% (at this point, a *cost of capital* approach is also being considered in life insurance). They would also ought to have at their disposal such a level of equity capital allowing them not to go bankrupt, in a year's time, with a very high probability (a priori 99,5%). Let us recall that the current scheme let the reserves' amount mostly assessed on deterministic sensible methods (conservative methods and hypotheses to estimate incurred losses, without any updating, or using a definitely wise rate). Regarding the solvency capital requirement (SCR), it follows the solvency margin system: the insurers are bound to carry an equity capital standard and latent surplus on financial investment high enough to compete with a solvency margin requirement, expressed as a percentage of the mathematical reserves in life insurance, and as a percentage of the earned premiums or of the incurred losses in non life insurance. This system is entirely deterministic and explicitly refers to the risk really taken on by the company. In particular, two firms with different risk profiles but the same accounting items are bound to the same level of capital[2]. We still ought to make it clear that, from a prudential point of view, this system appears to be quite efficient compared to the few insurance company failures that occurred in Europe during the last years.

The two new quantitative requirements framed by Solvency II, both refer to a VaR, but their nature strongly differs. The requirement on the technical reserves will not raise any major problem. Indeed, the data observed and used by the insurers are numerous and their stochastic methods to calculate the provisions are robust. Moreover, the percentile that has to be estimated is not very high, and thus the study focuses on the core of the distribution where the information is reliable. The assessment of the appropriate level of capital or Solvency Capital Requirement (SCR) is not that simple. In this case, the insurers do not directly observe the interest variable (the profit). The statistical environment needed to estimate the VaR is thus not directly available.

A standard[3] formula will be put forward aiming at an approach of the ruin probability criterion as accurate as possible. All the while, the project plans for the firms to be able to put together in-house patterns based on the modelling of all the variables that influence the solvency. Such models will allow to measure the capital standard needed by the company today in order not to go bankrupt in a year's time, with a probability of 99.5%. The construction of such a model is

---

[1] The liabilities of an insurance company are mostly represented by the technical reserves on one hand, and by the equity capital on the other.

[2] This statement ought to be lowered by the intervention of the regulatory authority (ACAM) which can enhance the capital requirement of the insurers.

[3] Within the framework of the Quantitative Impact Study QIS2 on the capital requirement on which may lead the Solvency II project, the Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS) published a model aiming at measuring each risk and aggregate the corresponding capital using a formula similar to the one from the American RBC.

a full problematic: the modelling of the base variables and their interactions must be precise. A model like this one also has to deal with the technical constraints linked with the computing power. Otherwise, the amount to be estimated is an extreme percentile of the distribution. A naive but robust empirical estimation using the empirical 99.5% percentile would require an extremely high number of simulations, each of these realizations resulting from the simulation of different phenomena such as the losses, the run-off of the technical reserves, the evolution of the financial credits … The estimation of such a percentile will require the use of some techniques deriving from the extreme value theory, which was specifically developed since the mid 70's and the work of Pickands (1975) and Hill (1975), and more recently of Smith (1987), Dekkers and Haan (1989), or Dekkers and al. (1989). It was soon practised to the financial and insurance problematics (Embrechts and al. (1997)). Yet, these results make use of tail data, which poses the new problem of the robustness of these data which, let us recall, are not observed but simulated. In fact, the internal models will mainly use a parametric modelling of the base variables (for instance, the evolution of the share prices) although these models do not exactly fit to the observations. In particular, the parameters are estimated on the entire distribution, and thus might not represent the tail distribution properly. In our problematic, the tail distribution will in fact engender the worst situations in terms of solvency, that is to say on the extreme values on which will be built the estimate of the 99.5% VaR.

This raises the problem of the robustness of this 99.5% VaR criterion, favouring the insured's security: the insurance firms might be tempted to vary the parameters and the base variable models, thus underestimating the SCR. First describing the specificities of the VaR computation in the realm of insurance, we present and illustrate the limits of the risk measurement suggested to assess the SCR in the Solvency II plan. In order to be clear, the main results of the extreme value theory are expounded in the appendix.

1. VaR computation in insurance

The adaptation of the VaR (and the TVaR) to the insurance problematics is a recent matter and compels to an approach totally different from the banking approach. In fact, the reference situation of the banking sector consists in estimating the VaR on an important sample of gains/losses on a portfolio or position. Massive data are available, with a high frequency. Historical VaR-type approaches are the base on which many banking models are built (see Christofferson and al. (2001)). The application of the banking techniques to the assets portfolio of an insurer requires some adjustments, mostly to take into account the fact that the funds are owned much longer. We might for instance refer to Feder and Morel (2006) who present a quantitative analysis of the subject. In the context of the assessment of the solvency capital, new difficulties occur; the nature of the data invalidates the historical approach. Here shall we recall the insurance cases in which the VaR are estimated; in practice. There are two that mainly appear, triggering different approaches:

- the provision computation using a 75% percentile of the loss distribution

- the SCR assessment to control a ruin probability, compelling the latter to be under 0.5% in a year's time.

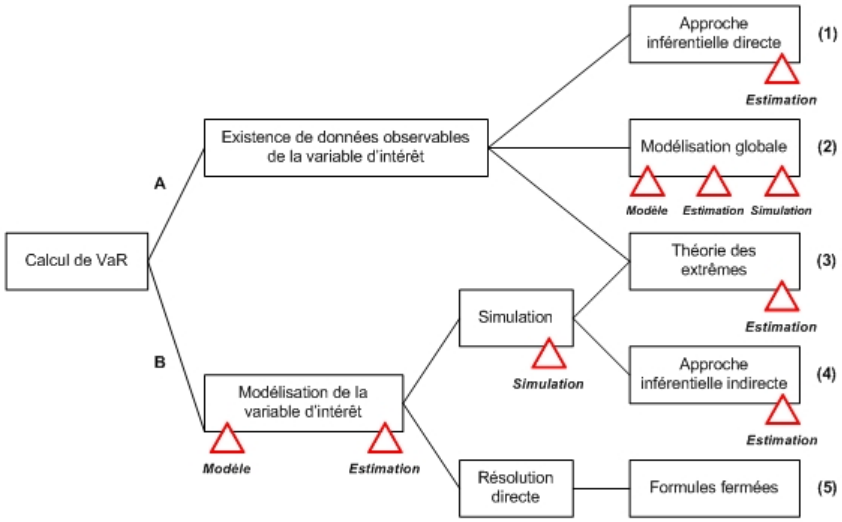
The first situation is a simple evolution of the actual prevailing one; indeed, the provisions are now calculated via a "best estimate" approach (see Blum and Otto (1998)). We agree that both computations (of an expectation and of a 75% VaR) are not fundamentally separated, on a statistical way. In both calculi, the data available allow the use of the classical inferential statistics techniques. Practically, the modelling of incurred losses using a parametric model is often suggested, leading to the estimation of the parameters of the underlying distribution (see Partrat and Besson (2005) for a recall of the most used distributions).

Things are completely different regarding the SCR assessment; only few data are available (at best, a few years of observation of the profit, for instance) and not any in the appropriate area of the distribution (if we had any, the insurer would have gone bankrupt). In the previous case, the interest variable was directly observed; here the VaR to be computed is now the result of a model, often complex, enlightening the different entries of the balance sheet of the insurer: the losses, the provisions, the financial assets…The reasoning will hence consist in building a model describing each item of the balance sheet; practically, the liability is modelled, the reserves being its main item, then the asset is modelled, and finally the potential interactions between the two. Having estimated the parameters of these models, simulation techniques make an estimate of the distribution of the profit available: the level of VaR wanted is then deduced from this estimation. Each stage of this reasoning entails risks that ought to be at least identified:

- a modelling risk: the model that is used provides an image of the reality that is not perfect; moreover, the models usually used to estimate both the assets and the liabilities tend to underestimate the extreme situations; this point might become a great disadvantage in the Solvency II approach, thus will we come back to it further;
- an estimation risk : the estimated parameters that feed the model are spoilt by an error. Its consequence might be severe when the model lacks robustness (see for instance Windcliff and Boyle (2004) for an example with the Markowitz model);
- a simulation risk: the profit's distribution will usually be estimated using a simulation and thus will only be approximate.

Furthermore, when it comes to the estimation of a percentile of high level (99.5% VaR), and knowing that the shape of the profit's distribution is generally hard to fit to a global parametric model, we ought to turn the extreme value techniques to finally calculate the risk measurement; making appear a new estimation risk on the tail distribution that is simulated. The figure 1 gives a synthesis of theses different risks and of the situations in which they occur.

Figure 1 – Typology of some of the risks encountered



The classical studies on the extreme variability of financial variables typically follow the A3 or A1 ramification of the diagram above; a high quantity of observations are available and the variables a re quite regular (they usually go up to the third or fourth moment). The estimation of a 75% VaR in insurance, aiming at computing Solvency provisions for instance, will at best correspond to the A1 ramification for the risk that provides a lot of observations (non life risks with a high frequency) or in most cases to the B4 or B5 ramifications (the VaR is usually given by a parametric model or a Monte Carlo method).

The SCR estimation derives from the B3 ramification which suffers from three risks: modelling, simulation and estimation (twice). The estimation and simulation risks are not specific to this work's problematic; the techniques controlling them are robust and reachable.

On the contrary, the modelling risk takes here a particular significance; in fact, and as we will illustrate later, the correct measure of the solvency capital compels to redo the modelling of the items of the balance sheet in order to take properly into account the extreme events and thus avoid an underestimation of the solvency capital.

2. Notations

Let us consider a n-sample of random variables (r.v.) $X_1,...,X_n$ independent and identically distributed (i.i.d.), F being their cumulative distribution function and $\overline{F}: x \mapsto 1 - F(x)$ their survival function. The order statistic associated $X_{n,n},\ldots,X_{1,n}$ will be defined by

$$\min\left\{X_1,\ldots,X_n\right\} = X_{n,n} \le X_{n-1,n} \le \ldots \le X_{1,n} = \max\left\{X_1,\ldots,X_n\right\}.$$

Otherwise, let F_u be the distribution function of the r.v. ${}_uX$ representing the surplus of X beyond the threshold u, when X exceeds the threshold u, that is

$$F_u(x) := \Pr\left[{}_uX \le x\right] = \Pr\left[X - u \le x \mid X > u\right].$$

x_F will be the upper bound of the space of outcomes of X, represented by the distribution function F, thus

$$x_F := \sup\left\{x \in \mathbb{R} \mid F(x) < 1\right\}.$$

And finally, N_u will be the number of observations exceeding the threshold u, thus

$$N_u = \sum_{i=1}^{n} 1_{X_i > u}.$$

3. Estimation of the extreme percentiles

We want to assess the amount that is exceeded with a low probability. Let us consider the standard situation where the realizations of the phenomenon are directly observed. The goal of this section is to present the different estimation methods that can be used and to measure the estimation errors made compared to the amount of data available.

The different methods are illustrated using a simulated sample of the realizations of a random variable of first type Pareto distribution.

3.1. Natural estimation

Since $X_{k,n}$ is a natural estimator of the first percentile $1-(k-1)/n$, a natural estimator of $F^{-1}(p)$ $(p \in]0;1[)$ is

$$([pn] - pn + 1)X_{n-[pn]-1,n} + (pn - [pn])X_{n-[pn],n},$$

where $[x]$ represents the integer part of x.

In practice, to be efficient, this method needs an important amount of data that is never available in this context.

3.2. Adjustment to a parametric distribution

A natural method consists in adjusting all the data to a parametric distribution, and then to estimate the percentile function with the level of probability wanted. In fact, let us recall that if $\hat{\theta}$ is the maximum likelihood estimator (m.l.e.) of the parameter θ of the distribution of X, and then $F_{\hat{\theta}}^{-1}(p)$ is the m.l.e. of $F_{\theta}^{-1}(p)$. Moreover, the BC bootstrap percentile methods (BCa) (see Efron and Tibshirani (1993) and the 4th paragraph) give confidence intervals of the estimation.

This estimation method is divided in two stages:

- Adjustment to a parametric distribution: choice between one or more distributions, estimation of the parameters, adequacy tests;
- Inverse of the distribution function (analytically when possible, numerically otherwise).

The main difficulty of this method consists in the choice of the parametric distributions that will be used. They follow a double constraint: they have to allow a numerical evaluation of their percentile function at any level wanted, and they must match the data. The latter point raises most of the difficulty since the estimation of the parameters is assessed on all of the observed distribution and seldom gives a good representation of the extreme values. Unless being sure that the chosen distribution for the adjustment is the real one from which the data are issued, which is in practice exceptional, the interest of this method is very limited in a context of estimation of extreme percentiles.

3.3. GPD Approximation

Known as the *Peaks Over Threshold* (POT) method, this technique is based on the Pickands-Balkema-de Haari theorem (see appendix B.3) which establishes that, when u is high enough, F_u can be approximated with a generalised Pareto distribution (GPD). Once the parameters of the GPD estimated (see appendix A3), since $\overline{F}(x) = \overline{F}(u) \times \overline{F}_u(x-u)$, the following estimate stands:

$$\overline{F}(x) \approx \frac{N_u}{n}\left(1+\frac{\hat{\xi}}{\hat{\beta}}(x-u)\right)^{-1/\hat{\xi}},$$

when $x > u$. The inverse of this distribution function provides us an estimator for the percentile function F^{-1} at high levels:

$$\hat{x}_p = u + \frac{\hat{\xi}}{\hat{\beta}}\left(\left(\frac{n}{N_u}(1-p)\right)^{-\hat{\xi}} - 1\right).$$

N.B. If the threshold chosen is one of the observations $X_{k+1,n}$ (for instance the empirical 99% percentile to estimate the 99.5% percentile) then $N_u = k$ and the estimator can be written as such:

$$\hat{x}_p = X_{k+1,n} + \frac{\hat{\xi}}{\hat{\beta}}\left(\left(\frac{n}{k}(1-p)\right)^{-\hat{\xi}} - 1\right),$$

when $p > 1 - k/n$.

In practice, this method raises a dilemma when it comes to choosing the threshold *u*. Indeed, *u* must be high enough for the GPD approximation to be right but not too close to the percentile that has to be estimated. Hence, enough statistical means would be available so that the estimation (using N_u) of the probability of being between *u* and this percentile is reliable. The problematic arisen by the choice of u is similar to the one of the choice of the number of observations to be used in the Hill estimation of the thickness of the tail distribution (see appendix C.2.4). A practical approach aiming at choosing the

threshold u consists in drawing the following function, the empirical estimator of the residual expectation of X

$$e_n(u) := \frac{1}{N_u}\sum_{i=1}^{n}\left[X_i - u\right]^+ ,$$

and to chose u so that $e_n(x)$ is almost linear when $x \geq u$. Indeed, because of the linearity of the residual expectation function of a GPD whose parameter is $\xi < 1$:

$$e(v) := \mathrm{E}\left[{}_v X\right] = \frac{\beta + \xi v}{1-\xi},$$

the wanted u will be as small as possible under the constraint that the empirical estimator of the residual expectation of the excess over u is almost linear.

One must not expect the right value for u from this method, although it gives a helpful hand. In practice, several values for u have to be tested.

3.4. Hill estimator

Let us consider the distribution function part of Frechet's minimum attraction domain (MAD) with $(\xi > 0)$. The tauberian theorem (see appendix B.2) shows that $\overline{F}(x) = x^{-1/\xi}\mathcal{L}_F(x)$, where $\mathcal{L}_F$ is a a slow variation function. Hence,

$$\frac{\overline{F}(x)}{\overline{F}\left(X_{k+1,n}\right)} = \frac{\mathcal{L}_F(x)}{\mathcal{L}_F\left(X_{k+1,n}\right)}\left(\frac{x}{X_{k+1,n}}\right)^{-1/\xi}.$$

When k is small enough and $x > X_{k+1,n}$, we can put aside the first-order ratio of the slow variation functions, then comes

$$\overline{F}(x) \approx \overline{F}\left(X_{k+1,n}\right)\left(\frac{x}{X_{k+1,n}}\right)^{-1/\xi}.$$

From which we deduce F's estimator:

$$\hat{F}(x) = 1 - \frac{k}{n}\left(\frac{x}{X_{k+1,n}}\right)^{-1/\xi_k^H},$$

when $x > X_{k+1,n}$. The inverse of this function gives the Hill estimator of the percentile function for high levels $(p > 1 - k/n)$:

$$\hat{x}_p^H = X_{k+1,n}\left(\frac{n}{k}(1-p)\right)^{-\xi_p^H}.$$

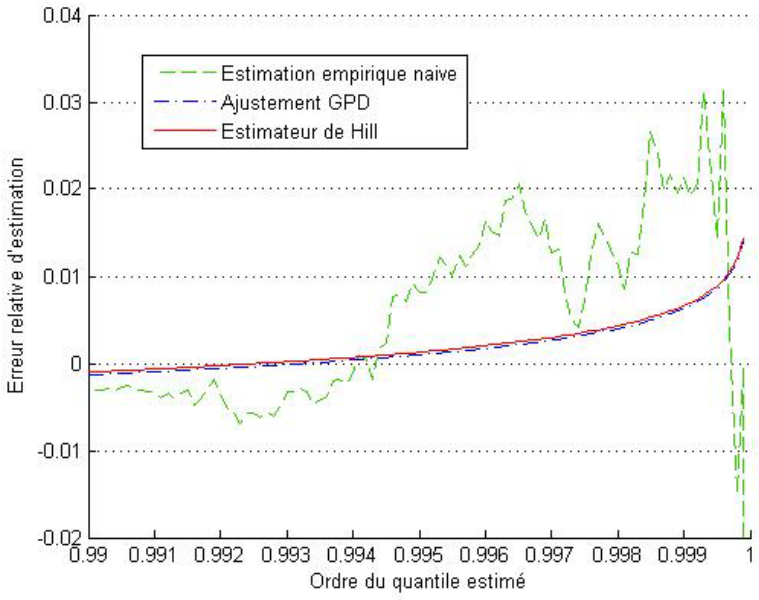
The dilemma of k's choice is similar to the one encountered in estimating the tail parameter ξ (see appendix C.2.4).

3.5. Illustration

The **Erreur ! Source du renvoi introuvable.** shows the different techniques used to estimate a high percentile of a first type Pareto distribution, whose distribution function is

$$F(x) = 1 - \left(\frac{x_0}{x}\right)^{\alpha}, \text{ when } x > x_0 .$$

Figure 1 - Estimation of an extreme percentile: relative estimation error



Such a distribution belongs to the Frechet's MAD. The POT method (which is right with a first type Pareto distribution) and the Hill estimator give similar results that are much more accurate than the natural estimator.

4. Application of the bootstrap

The bootstrap method, first introduced by Efron (1979), is commonly used in insurance (see Partrat and Jal (2002) and Verall and England (1999)). Anyone interested in more general purpose of the bootstrap method can find further information on its applications to econometrics in Horowitz (1997).

When it comes to estimating the extreme values, its application enables us to establish the robustness of the estimations by providing useful confidence intervals.

4.1. Presentation

The principle of the classical method consists in noticing that, for a big enough sample, the distribution function F of the underlying distribution can be approached with the empirical distribution function:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_n \le x\}} .$$

In practice, this general principle makes us derive from the Bootstrap method by adapting it to the different contexts: time series, linear models, non linear regression models … Further is exposed the use of the method in the simple case of building a confidence interval for a central estimator.

To estimate the function $I(g)=\int g(x)dF(x)$ from F, the classical simulation approach brings us to calculate

$$I_n(g)=\int g(x)dF_n(x)=\frac{1}{n}\sum_{i=1}^{n}g(X_i),$$

which is an approximation of $I(g)$. The asymptotical theory provides information on the distribution of the statistic $I_n(g)$: The central limit theorem helps us to prove that the asymptotical distribution is Gaussian and hence to calculate confidence intervals such as

$$J_\alpha=\left[I_n(g)-\frac{\sigma_g}{\sqrt{n}}\varphi^{-1}\left(1-\frac{\alpha}{2}\right),I_n(g)+\frac{\sigma_g}{\sqrt{n}}\varphi^{-1}\left(1-\frac{\alpha}{2}\right)\right],$$

where $\sigma_g^2=\mathrm{Var}\left(g(X)\right)$ has to be estimated. However, in some situations, the asymptotical approximation is not enough accurate, and the bootstrap method gives an alternative to find information on the distribution of $I_n(g)$.

This technique consists in using the realizations of $I_n^b(g)$, when $b=1,\ldots,B$ and $B\le n^n$ that are given by inverting the initial sample $(X_1,\ldots,X_n)$ with the « bootstraped » samples $\left(X_1^b,\ldots,X_n^b\right)$ which are drawn (with replacement) in $(X_1,\ldots,X_n)$. B is known as the length of the bootstrap sample.

Indeed, using simulation to value statistics consists in generating samples with the empirical distribution. Though, drawing in the empirical distribution is simply given by a draw with replacement of the *n* values in the initial sample. At least n^n « bootstraped » samples are hence found, from which we will calculate the empirical estimators of the interest statistics.

Thus, by a random draw in the original sample, the bootstrap can interfere in it so as to obtain new estimators to the parameters.

Once *B* estimators are available $I_n^b(g)$ of $I(g)$, the equations

$$\hat{\mu}\left[I_n^1(g),\ldots,I_n^B(g)\right]=\frac{1}{B}\sum_{b=1}^{B}I_n^b(g),$$

and

$$\hat{\sigma}^2\left[I_n^1(g),\ldots,I_n^B(g)\right]=\frac{1}{B-1}\sum_{b=1}^{B}\left[I_n^b(g)-\hat{\mu}\right]^2,$$

give a confidence interval such as

$$J_\alpha=\left[\hat{\mu}-\frac{\hat{\sigma}}{\sqrt{B}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right),\hat{\mu}+\frac{\hat{\sigma}}{\sqrt{B}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right].$$

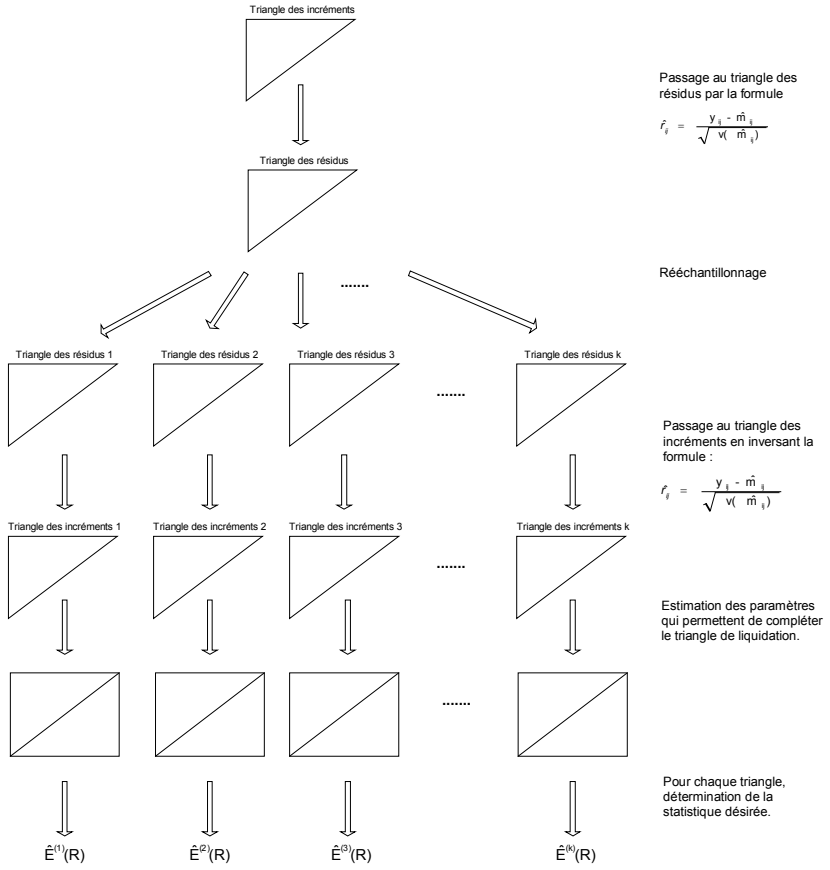
This technique combined with to parallel techniques enabling to narrow the results: the percentile bootstrap and the BCa method. These techniques are described infra in the particular case of the estimation of percentiles.

However, the bootstrap method is not relevant in all cases: for example, it does not work when it comes to looking for the maximum value of the sample. Indeed, this building method does not provide samples with values superior to the maximum of the initial sample. For this type of problem, other methods exist, such as the one introduced by Zelterman (1993) consisting in re-sampling the intervals between two following values of the initial data instead of re-sampling the data themselves. This semiparametric method is illustrated is the case of estimating with intervals the maximum intensity of earthquakes in different areas by Bottard (1996).

The usual bootstrap is although very efficient for all the mean-based statistics.

The bootstrap technique is commonly used in insurance to study the variability of the costs of the losses, and to find predicting errors for different reserving methods, in particular for the Chain Ladder based methods and the generalised linear models (see Partrat and Jal (2002)). The general scheme (taken from Planchet, Thérond and al. (2005)) is simplified with the Figure 2.

Figure 2 – Bootstrap method in non-life reserving



4.2. Calculus of a confidence interval for a VaR

The different bootstrap techniques are exposed further in this paragraph in the specific context of the estimation of a percentile using intervals.

In the classical case, where the observations are supposed to be realizations of random variables whose underlying distribution is parametric (Gaussian, log-normal, Pareto, Benktander, etc.), the estimation of a percentile of any level can follow this method:

- Estimation of the parameter θ (using the maximum likelihood when possible, or with another method): $\hat{\theta}$;
- Inverse of the distribution function;
- Particular estimation of the VaR: $\hat{VaR}(q) = F_{\hat{\theta}}^{-1}(q)$;
- Search for a confidence interval.

The estimator thus calculated has to be accurate. In the particular case where the parameter is estimated with maximum likelihood estimator (m.l.e.), the general properties of the estimator make of $\hat{VaR}(q)$ the m.l.e. of the VaR (as a function of a m.l.e.). Thus, it is asymptotically unbiased and Gaussian. An estimate of its asymptotical variance will enable us to build a confidence interval.

The distribution of the statistic $\hat{VaR}(q)$ is however hard to establish. It is then advised to use the bootstrap method.

To estimate a VaR, the initial sample is made of the n observations of X used to estimate the parameters of the model. The interest statistic is:

$$\hat{VaR}(q) = F_{\hat{\theta}}^{-1}(q).$$

The confidence interval we look for is like:

$$\Pr\left(b_1 \le VaR(q) \le b_2\right) = 1 - \alpha.$$

Three bootstrap techniques can give the appropriate confidence interval:

- Classical bootstrap: estimation of the bootstrapped variance of $\hat{VaR}(q)$ that is further used with Gaussian asymptotical hypothesis;
- Percentile bootstrap: ranking of the particular estimations that are given for each bootstrap sample and establishing of the bounds of the confidence interval with the corresponding estimations;
- Direct estimation of the bootstrapped confidence interval with the BCa method (Bias corrected and accelerated).

In the first approach, the estimation of the bootstrapped variance simply consists in computing the empirical variance of the sample of the $\hat{VaR}_b(q) = F_{\hat{\theta}_b}^{-1}(q)$ where $\hat{\theta}_b$ is the parameter estimated on the b-th bootstrap sample. A confidence interval is hence easily given knowing that the maximum likelihood estimator is asymptotically Gaussian. In practice, this method is not recommended for the non-parametric problems (for which the estimation of the percentile of each sample would rest on its empirical estimator).

The percentile bootstrap gives a better alternative for non-parametric problems and is much more efficient. This technique consists in ranking the $\hat{VaR}_b(q)$ in increasing order and to take the $B \times \alpha/2$ and $B \times (1-\alpha/2)$-th's highest values as bounds for the confidence interval. This technique fits in the non parametric problems because it does not use the asymptotic properties of the m.l.e.

Like the percentile method, the BCa method (see Efron (1987)) offers to optimize the process of building the confidence interval without using the asymptotical properties of the m.l.e. This method consists in determining the confidence bounds taking the order value $B \times \beta_i$ of the sample of $\hat{VaR}_b(q)$ for b_i, where

$$\beta_1 = \Phi\left(z_0 + \frac{z_0 + u_{\alpha/2}}{1 - \gamma\left(z_0 + u_{\alpha/2}\right)}\right) \text{ and } \beta_2 = \Phi\left(z_0 + \frac{z_0 + u_{1-\alpha/2}}{1 - \gamma\left(z_0 + u_{1-\alpha/2}\right)}\right),$$

and Φ the distribution function of the standard normal distribution, u_c the c-percentile of the same distribution, the correcting bias parameter $z_0 = \Phi^{-1}(k)$

where k is the proportion of the bootstrapped samples for which the estimated VaR is lower than the VaR estimated on all the samples. Finally, the speeding parameter γ is estimated with:

$$\gamma = \frac{\sum_{i=1}^{n}\left(\widetilde{VaR}_i - \widetilde{VaR}\right)^3}{6\left[\sum_{i=1}^{n}\left(\widetilde{VaR}_i - \widetilde{VaR}\right)^2\right]^{3/2}},$$

where $\widetilde{VaR}_i$ is the estimate of the VaR deriving from the n° i Jackknife sample (observed sample from which the i-th observation has been removed) and $\widetilde{VaR}$ is the mean of these n estimates (see Quenouille (1949)).

4.3. Numerical illustration

These methods are numerically illustrated in the case of a log-normal distribution. In this case, the following equation is immediately given:

$$VaR_p(X) = F^{-1}(p) = \exp\left(m + \sigma\varphi^{-1}(p)\right),$$

and thus the p-th maximum likelihood estimator of the VaR:

$$\hat{VaR}_p(X) = \exp\left(\hat{m} + \hat{\sigma}\varphi^{-1}(p)\right),$$

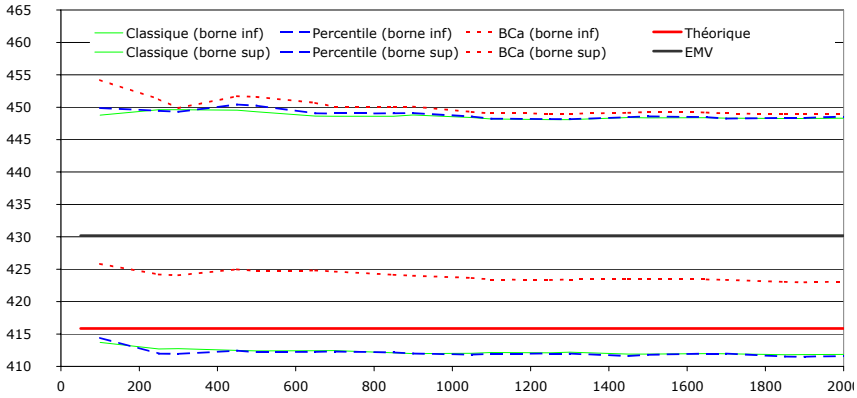
where $\hat{m} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \hat{m}\right)^2}$.

The following numerical values are taken: $m = 5$, $\sigma = 0,4$, which lead to a theoretical 99.5% VaR of:

$$VaR_{99,5\,\%}(X) = \exp\left(m + \sigma\varphi^{-1}(p)\right) = 415,85.$$

This percentile is estimated on a 1000-length sample and the 90% confidence interval is determined with the classical, percentile and BCa method. The following typical results are hence given according to the length of the sample:

Figure 3 – Confidence intervals (99.5% VaR)

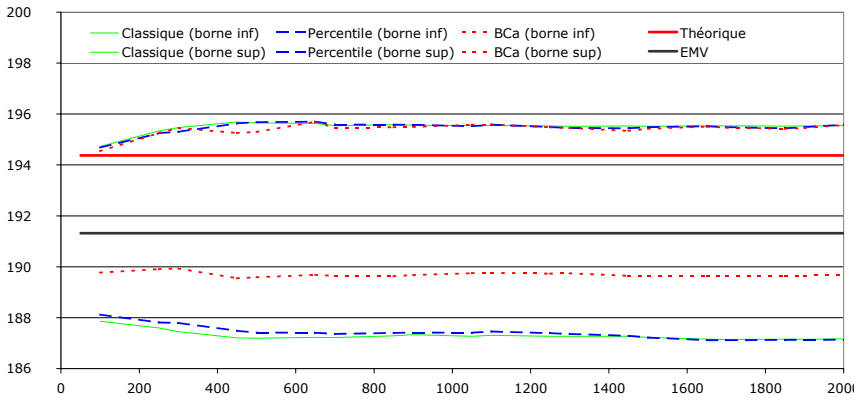


The classical and percentile methods are quite close. The BCa method provides an interval whose upper bound is close to the one provided by the two

other methods, but the lower bound is way higher. In the previous example, we notice that when the real value is in the confidence interval of both the "classical" and the "percentile" methods, it is smaller than the lower bound of the confidence interval given by the BCa method.

The results of the bootstrap methods are although not very robust for a percentile of high level. For a lower percentile, as the 75% percentile, the estimations are much more reliable:

Figure 4 – Confidence intervals (75% VaR)



5. Robustness of the SCR

This paragraph aims at specifying the limits of an extreme VaR-type criterion to establish the solvency capital of an insurance company. Indeed, the complexity of the process that has to be settled lowers the impact of the results of an internal model. Two aspects are specifically developed and illustrated with a simple example: the estimation error of the base variables parameters and the specification error of the model of these variables.

5.1. Estimation of the base variables parameters

Let us consider the simplified internal model inspired by the works of Deelstra and Janssen (1998) in which the losses of the year are modelled with a random variable with a log-normal distribution and paid at the end of the year. Let us suppose that the return on investment on the period is a Gaussian variable, independent from the losses. Let $B \sim \mathcal{LN}(m,s)$ denote the r.v. modelling the amount to be paid at the end of the period and $\rho \sim \mathcal{N}(\mu,\sigma)$ denote the return on investment on the period.

Let us find the minimum amount of asset a_0 that the insurer must have at his disposal at the beginning of the period so as not to go bankrupt at the end of the period, with a probability higher than $1-\alpha$. Formally, a_0 is the solution to the following optimisation problem:

$$a_0 = \min\left\{a > 0 \mid \Pr\left[ae^{\rho} \geq B\right] > 1-\alpha\right\}.$$

As $Be^{-\rho} \sim \mathcal{LN}\left(m-\mu, \sqrt{s^2+\sigma^2}\right)$, this analytic value of a_0 is given:

$$a_0 = \exp\left\{m-\mu+\Phi^{-1}(1-\alpha)\sqrt{s^2+\sigma^2}\right\}.$$

The sensibility of a_0 to the parameters of the basic models is then studied. For instance, the liabilities are given by:

$$\frac{1}{a_0}\frac{\partial a_0}{\partial m} = 1,$$

and

$$\frac{1}{a_0}\frac{\partial a_0}{\partial s} = \frac{\Phi^{-1}(1-\alpha)}{\sqrt{1+\sigma^2/s^2}}.$$

A relative estimation error of the parameter *m* leads to the same relative error on a_0. Furthermore, a relative error of 1% on *s* leads to a relative error of $\Phi^{-1}(1-\alpha)\Big/\sqrt{1+\sigma^2/s^2}$ on a_0. With a 99.5% VaR, when $s \simeq \sigma$, the error is 1.82 times higher.

5.2. Simulation

The complexity joint to the modelling of the profit of an insuring activity makes it inevitable to resort to the simulation techniques to have numerical results. Although the basic principle of these methods is simple and universal, (using the strong convergence of the law of large numbers), an efficient implementation requires being at least cautious on some points. Indeed, the use of simulated results engenders different kinds of error:

- fluctuations of sampling linked to the finite numbers of draw that are done;
- discretization bias when transforming a continuous model in its discrete versions;
- errors associated to the approximations used by some of the inversing techniques;
- bias incurred by an inappropriate choice of the generator of random numbers.

Moreover, the algorithms used must allow a quantitative control of all these error sources in order to be able to grade the number of draw necessary for the level of accuracy wanted (a confidence level is decided *a priori*).

5.3. Specification of the model

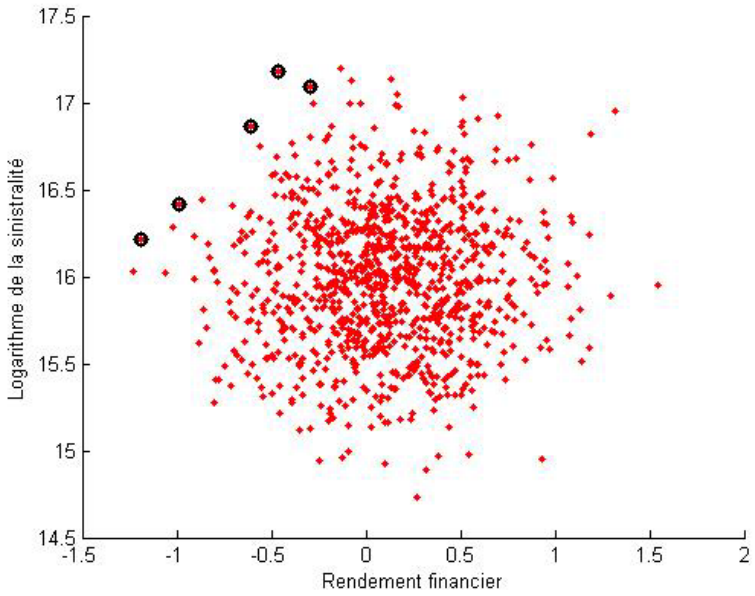
5.3.1. Context and motivation

The building of an internal model requires the modelling of the different variables that influence the solvency of the insurer. In order to do so, the first stage consists in the modelling of these different variables. For parametric adjustments, we naturally estimate the parameters and the adequacy tests on all the data available. However, in this problematic, the tail distributions are the

one that have an influence on the level of the SCR required. Although, these are often insufficiently described by the global approach, using a simple and unique parametric distribution. In particular, most of the usual models (log-normal, Benktander, etc.) lead to distributions in the Gumbel's MAD, whereas the observations tend to have tail distributions that are in the Fréchet's MAD. Indeed, the stochastic models that currently work in the insurance companies have initially been created to study the profitability or to calculate technical reserves. Knowing that the results are most of the time evaluated with expectation-variance type criteria (or VaR in the core distribution – 75% for instance), the modelling of extreme events has its own importance in this context. The Solvency II approach enlightens only the extreme values of the interest variable: the capital that has to be available today so as not be ruined in a year's time. This remark has to be lowered by the fact that most of the insurers that invest in the development of internal models expect (except for a few adjustments) that this model enables them to refine their strategic plan and to answer to the future accounting requirements resulting from the Phase II of the insurance contracts standard.

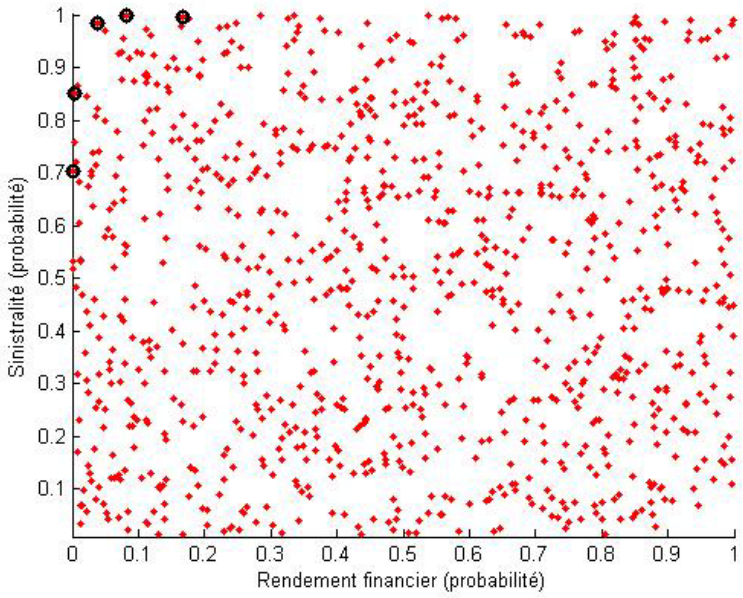
Let us consider the model we just exposed. The Figure 5 shows a 1000-realization sample of the amount of asset that a society must have in 0 so as not be ruined in 1.

Figure 5 – Simplified internal model: identifying the extreme values



The five underlined spots correspond to the five scenarios that lead to maximum values.

Figure 6 – Simplified internal model: identifying the extreme values (probabilities)

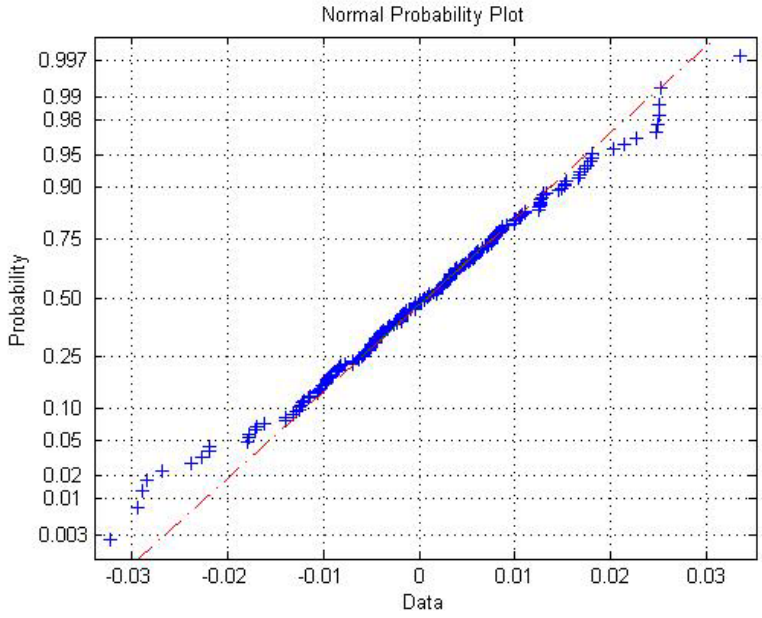


The Figure 6 represents the probability linked to each variable. We can notice that these spots are found in the tail distribution of at least one of the base variables. It is precisely in the tail of the distribution that the fit is the more imprecise.

5.3.2. Advanced modelling

Let us consider the Figure 7 that exposes the normal QQ-plot of the daily return on the TOTAL security (from July 2005 to April 2006).

Figure 7 – Daily return on the TOTAL security: QQ-plot of the empirical distribution against the normal distribution



Graphically, we can already notice that the adjustment is globally satisfying but that the tail distributions of the model (in particular the one for the negative return that is important for us) are too thin. Indeed, the statistical tests of Jarque-Béra and Lilliefors with a 5% threshold lead to the non-rejection of the hypothesis of normality of the return. However, in the case of determining the SCR based on a 99.5% VaR criterion, the Gaussian model would tend to minimize the risk taken when investing on this share. For instance with the data of the TOTAL security, the 0.5% percentile of the daily return observed is −0,0309 whereas the same percentile for the adjusted Gaussian model is −0,0286, which represent an error approxiamting 7,5 %. To remedy this, in the situations where we have an important number of observations available, we could be tempted to adopt a non-parametric approach using the empirical distribution. This is although not satisfying in practice, since these approaches are technically harder to implement, require longer simulation time, and go against an easy comprehension of the model. In this context, it is natural to keep a parametric approach and to look for a model that best represents the tail distributions.

The following model is a natural extension of the Gaussian return.

Suppose that the return of the asset on the period follows the process:

$$\rho = \mu + \sigma_0 \varepsilon_0 + \sigma_u \sum_{i=1}^{N} \varepsilon_i \ ,$$

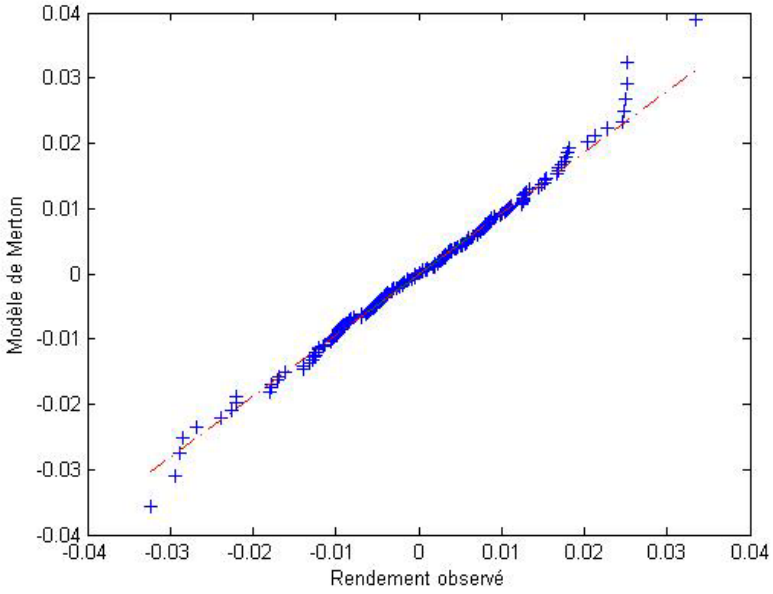
where $\varepsilon_0, \varepsilon_1, \ldots$ are i.i.d.r.v of $\mathcal{N}(0;1)$ distribution and independent of $N \sim \mathcal{P}(\lambda)$ $(\lambda > 0)$. This process is a mono-periodic version of the Merton process (1976). The maximum likelihood estimation requires the use of numerical techniques (see Planchet and Thérond (2005b)).

In our problematic, we can still try to number the parameters by hand so as to weight the tail distributions. For instance, when globally settling the expectation and the variance of the return (identical to the one of the Gaussian model), we can weight the variability represented by the jumping component so that the tail distribution of the model is as heavy as the one of the observations. Formally, it is like choosing λ, σ_0 and σ_u under the constraint:

$$\sigma_0^2 + \lambda\sigma_u^2 = \sigma^2 .$$

The Figure 8 represents this adjustment.

Figure 8 – Daily return of the TOTAL security: QQ-plot empirical distribution against Merton model



Compared to the Figure 7, we can observe that the tail distributions have been weighted. Generally, this reasoning shown for the modelling of a financial investment will have to be looked at for all the base variables of the asset and the liabilities in order not to underestimate the SCR.

6. Conclusion

The Solvency II project aims at establishing a prudential framework standardized and consistent to all the insurance societies in Europe. In that purpose, it plans in particular to incite the firms to have a better modelling of their risks by allowing them to build internal models leading to a solvency capital that is lower than the one given by the « standard formula ». However, we have witnessed the relative robustness of the 99.5% VaR criterion on variables that are not observed but generated by the internal model. In particular, the stochastic models currently working in the insurance companies

(for both the assets and the liabilities), are not prepared for the estimation of the extreme values and are most of the time built from a modelling of the « base variables » which under estimate the tail distributions. The use of such models as they are at first might lead to an underestimation of the SCR. That's why the insurance companies who want to develop an internal model must model the tail distributions more precisely. In the meantime, the authorities that will publish the internal models will have to be particularly careful to these during the validation process. Especially, a European standardisation of these processes is inevitable for fear of integrating a competition distortion between the different participants. Globally, when the control criterion of a ruin probability which underlay the Solvency II project is taken into account, the models usually used in the insurance firms have to be totally rethought.

Appendix A: Generalised Pareto Distribution (GPD)

A.1 Definition

Definition 1 : Generalised Pareto distribution

Lett $H_{\xi,\beta}$ *be the distribution function defined for* $\beta > 0$ *with*

$$H_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\beta} x\right)^{-1/\xi}, & \text{si } \xi \neq 0, \\ 1 - \exp(-x/\beta), & \text{si } \xi = 0. \end{cases}$$

This distribution function corresponds to the Generalised Pareto Distribution (GPD) with the ξ *and* β *parameters. It is defined when* $x > 0$ *if* $\xi > 0$ *and when* $0 \leq x \leq -\beta/\xi$ *if* $\xi < 0$.

Further we will denote as $D(\xi,\beta)$ the domain of the function $H_{\xi,\beta}$.

It has been showed (see Denuit and Charpentier (2005)) that the Generalised Pareto Distribution can be seen as a log-gamma or, when $\xi > 0$, as a mix of exponential distributions whose parameter follows a gamma distribution.

A.2 A few properties

The following results are given for *Y* a random variable distributed after a GPD of parameters (ξ,β).

Property 1. If $\xi < 1$, then

$$\mathrm{E}\left(1 + \frac{\xi}{\beta} Y\right)^{-r} = \frac{1}{1+\xi}, \text{ when } r > -1/\xi,$$

$$\mathrm{E}\left[\ln\left(1 + \frac{\xi}{\beta} Y\right)\right]^{k} = \xi^k k!, \text{ when } k \in \mathbb{N},$$

$$\mathrm{E}\left[Y\left(\bar{H}_{\xi,\beta}(Y)\right)^{r}\right] = \frac{\beta}{(r+1-\xi)(r+1)}, \text{ when } (r+1)/\xi > 0.$$

Property 2. The random variable *Y* has moments of orders up to $\lfloor \xi^{-1} \rfloor$ and

$$\mathrm{E}\lfloor Y^r \rfloor = \frac{\beta^r \Gamma(\xi^{-1} - r)}{\xi^{r+1} \Gamma(\xi^{-1} + 1)}, \text{ when } r \leq \lfloor \xi^{-1} \rfloor.$$

Property 3. (Stability) The random variable $Y_u = [Y - u \mid Y > u]$ follows a GPD of parameter $(\xi, \beta + \xi u)$. We deduce that if $\xi < 1$, then for any $u < y_F$,

$$\mathrm{E}[Y - u \mid Y > u] = \frac{\beta + \xi u}{1 - \xi}, \text{ when } \beta + u\xi > 0.$$

We recall that y_F is the upper bound of the space of outcomes of *Y*, so

$$y_F = \sup\{y \in \mathbb{R}, F(y) < 1\}.$$

A.3 Estimation of the parameters

Let us consider a *n*-sample $(Y_1,\ldots,Y_n)$ of the random variable Y whose distribution function is $H_{\xi,\beta}$.

A.3.1 Maximum likelihood method

The density function f of Y is

$$f(y)=\frac{1}{\beta}\left(1+\frac{\xi}{\beta}y\right)^{-1/\xi-1}, \text{ when } y\in D(\xi,\beta).$$

We deduce the log-likelihood

$$\ln L(\xi,\beta;Y_1,\ldots,Y_n)=-n\ln\beta-\left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\ln\left(1+\frac{\xi}{\beta}Y_i\right).$$

Using the parameter $\tau=\xi/\beta$, and equalling zero the partial derivatives of the log-likelihood lead to the system

$$\begin{cases}\xi=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\ln(1+\tau Y_i)=:\hat{\xi}(\tau),\\ \dfrac{1}{\tau}=\dfrac{1}{n}\left(\dfrac{1}{\xi}+1\right)\displaystyle\sum_{i=1}^{n}\frac{Y_i}{1+\tau Y_i}.\end{cases}$$

The maximum likelihood estimator of (ξ,τ) is $\left(\hat{\xi}=\hat{\xi}(\hat{\tau}),\hat{\tau}\right)$ where $\hat{\tau}$ is the solution of

$$\frac{1}{\tau}=\frac{1}{n}\left(\frac{1}{\hat{\xi}(\tau)}+1\right)\sum_{i=1}^{n}\frac{Y_i}{1+\tau Y_i}.$$

This last equation can be numerically solved by an iterative method as long as we have an initial value τ_0 not too far from τ. In practice, this initial value can be found with the method of moments (as long as these exist to the second order) or with the method of percentiles.

When $\xi>-1/2$, Hosking and Wallis (1987) have exposed that the maximum likelihood estimators are asymptotically normal:

$$n^{1/2}\left(\hat{\xi}_n-\xi,\frac{\hat{\beta}_n}{\beta}-1\right)\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}\left[0,(1+\xi)\begin{pmatrix}1+\xi & -1\\ -1 & 2\end{pmatrix}\right].$$

This result enables in particular to compute approximate estimation errors made by the maximum likelihood estimators.

A.3.2 Method of moments

The results of the paragraph A.2, when $\xi<1/2$, give the two first moments of Y:

$$\mu_1 := \mathrm{E}[Y] = \frac{\beta}{1-\xi},$$

and

$$\mu_2 := \mathrm{Var}[Y] = \frac{\beta^2}{(1-\xi)^2(1-2\xi)}.$$

We deduce that

$$\xi = \frac{1}{2} - \frac{\mu_1^2}{2\mu_2} \qquad \text{and} \qquad \beta = \frac{\mu_1}{2}\left(1 + \frac{\mu_1^2}{\mu_2}\right).$$

Putting the empirical estimators in place of μ_1 and μ_2 give the method of moments estimators $\hat{\xi}_{MM}$ and $\hat{\beta}_{MM}$. These estimators are easy to implement, but work only when $\xi < 1/2$, they require knowing this information *a priori*.

A.3.3 Method of probability-weighted moments

Hosking and Wallis (1987) introduced an interesting alternative to the method of moments. Noticing that

$$\omega_r := \mathrm{E}\left[Y\bar{H}^r_{\xi,\beta}(Y)\right] = \frac{\beta}{(r+1)(r+1-\xi)}, \text{ when } r = 0;1,$$

we find

$$\beta = \frac{2\omega_0\omega_1}{\omega_0 - 2\omega_1} \qquad \text{and} \qquad \xi = 2 - \frac{\omega_0}{\omega_0 - 2\omega_1}.$$

Putting the empirical estimators in place of ω_0 and ω_1, the estimators with the method of probability-weighted moments are given: $\hat{\xi}_{PWM}$ and $\hat{\beta}_{PWM}$.

With a lowered length sample, Hosking and Wallis (1987) exposed that, when $\xi < 1/2$, these estimators are more efficient than those of the maximum likelihood. The validity domain of this method is its main limit to a use for insurance: when most of the risks have two-ordered moments, it is not always as such (storms, earthquakes, industrial risks civil liability, etc.). Generalisations (Diebolt and al. (2005b)) allow extending this method to $\xi < 3/2$.

A.3.4 Method of percentiles

The percentile function of a GPD of parameter (ξ,β) is given by

$$q_p := H^{-1}_{\xi,\beta}(p) = \frac{\beta}{\xi}\left[(1-p)^{-\xi} - 1\right].$$

Noticing that

$$\frac{q_{p_2}}{q_{p_1}} = \frac{(1-p_2)^{-\xi} - 1}{(1-p_1)^{-\xi} - 1}, \text{ when } p_1, p_2 \in]0;1[,$$

The solution $\hat{\xi}_{MQ}$ of

$$\frac{Q_{p_2}}{Q_{p_1}} = \frac{\left(1-p_2\right)^{-\xi}-1}{\left(1-p_1\right)^{-\xi}-1}$$

where Q_{p_1} and Q_{p_2} are the empirical percentiles of level p_1 and p_2, is the percentile estimator of ξ.

A.3.5 Bayesian methods

Recent developments (see Coles and Powell (1996) or Diebolt and al (2005a)) suggest bayesian approaches to estimate the parameters of the GPD. Using the *Markov Chain Monte Carlo* (MCMC) algorithms, these methods enable us to integrate information *a priori* (experts' advice) in these contexts where we have a reduced number of observations at our disposal.

Appendix B: Probabilistic results

In this paragraph are successively reminded some probability calculi results about the distribution of the maximum, the thickness of the tail distributions and the excess over a threshold. The different results are demonstrated in Embrechts and al. (1997).

B.1 Distribution of the maximum

Definition 2 : Same type distributions

Two r.v. X and Y are said to be of same type *if two constants* $a \in \mathbb{R}$ *and* $b > 0$ *such as* $Y =_d a + bX$.

Definition 3 : Fisher-Tippett-Gnedenko

Let us consider a sequence $X_1, X_2, \ldots$ *of i.i.d.r.v. If a sequence of real numbers* $\left(a_n\right)$*, a sequence of positive numbers* $\left(b_n\right)$ *and a non-degenerated distribution H exist such as*

$$\frac{X_{1,n}-a_n}{b_n} \rightarrow_d H ,$$

Then H is of the same type as one of the following distributions:

$$G(x) = \exp\left\{-\left(1+\xi x\right)^{-1/\xi}\right\}, \quad \text{pour } 1+\xi x \geq 0,$$
$$\Lambda(x) = \exp\left\{-\exp\left(-x\right)\right\}, \quad \text{pour } x \in \mathbb{R}.$$

The distribution functions $G_+\left(\xi > 0\right)$, $\Lambda\left(\xi = 0\right)$ and $G_-\left(\xi < 0\right)$ respectively correspond to the Fréchet, Gumbel and Weibul distributions.

Definition 4 : Maximum Attraction Domain

A distribution function F that responds to the hypotheses of the Fisher-Tippett-Gnedenko theorem will be known has part of the maximum attraction domain (MAD) of H.

The Jenkinson-Von Mises representation provides a synthetic characterization of the extreme distributions: the Generalized Extreme Value (or GEV). The GEV distribution function has the following shape:

$$H_{\xi\mu,\sigma}(x) = \begin{cases} \exp\left\{-\left(1+\xi\dfrac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\} & \text{si } 1+\xi\dfrac{x-\mu}{\sigma} > 0, \xi \neq 0, \\ \exp\left\{-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right\} & \text{si } \xi = 0, \end{cases}$$

Where the localisation parameter μ is directly linked to the most probable value of the distribution; thus, it provides an approximate information on the core distribution whereas σ is the dispersion parameter which indicates the spread of the extreme values. Finally, ξ is the tail index previously introduced.

B.2 Thickness of the tail distributions

Definition 5 : Regularly varying function

A function $h : \mathrm{R}_+^* \to \mathrm{R}_+$ *is said to have a regular variation (in* $+\infty$*) with index* α *if* h *verifies*

$$\lim_{x\to+\infty} \frac{h(tx)}{h(x)} = t^{\alpha} \text{ for any } t > 0 .$$

If $\alpha = 0$, we will call it slow variation; if $\alpha = \infty$ fast variation.

Definition 6 : Tauberian theorem

A r.v. with a distribution function F and a Laplace transform L_F *is a regular variation with index* $-\alpha (\alpha \leq 0)$ *if the following equivalent conditions are true (*$\mathcal{L}_F$ *,* $\mathcal{L}_{F^{-1}}$ *,* $\mathcal{L}_L$ *and* $\mathcal{L}_f$ *stand for the slow variation functions):*

(i) $\overline{F}$ *is regularly varying with index* $-\alpha$ *, that is* $\overline{F}(x) = x^{-\alpha}\mathcal{L}_F(x)$.

(ii) The percentile function has regular variation:

$$F^{-1}(1-1/x) = x^{1/\alpha}\mathcal{L}_{F^{-1}}(x) .$$

(iii) The Laplace transform of F *verifies* $L_F(t) = t^{\alpha}\mathcal{L}_L(1/t)$.

(iv) If the density exists and verifies $xf(x)/\overline{F}(x) \to \alpha$ *when* $x \to +\infty$ *, then the density has regular variation of index* $-(1+\alpha)$ *, that is* $f(x) = x^{-(1+\alpha)}\mathcal{L}_f(x)$.

The condition related to the Laplace transform enables us to establish that the property of regular variation with a decided parameter is still valid after a convolution.

B.3 Distribution of the excess over a threshold

Definition 7 : Pickands-Balkema-de Haari

A distribution function F *belongs to the maximum attraction domain of* G_ξ *exclusively if a positive function* $\beta(.)$ *exists such as*

$$\lim_{u \to x} \sup_{x>0} \left\{ \left| {}_u F(x) - H_{\xi,\beta(u)}(x) \right| \right\} = 0 .$$

This theorem makes the link between the parameter of the maximum attraction domain distribution and the limit behaviour of the excess over a high threshold. Particularly, the tail index ξ equals the parameter of the GPD distribution that describes the residual cost of the losses over a high enough threshold.

In particular, this enables us to distinguish the thick-tailed distributions that belong to the Fréchet's MAD $(\xi > 0)$ from the thin-tailed distributions belonging to the Gumbel's MAD $(\xi = 0)$. The following chart provides the limit behaviour of some usual distributions.

Tableau 1 – Extreme behaviour of usual distributions in insurance

Thick-tailed distributions $\xi > 0$	**Thin-tailed distributions** $\xi = 0$	**Upper-bounded distributions** $\xi < 0$
Cauchy Pareto log-gamma Student α-stable $(\alpha < 2)$	Gamma Gaussian log-normal Weibull Benktander	uniform Beta

The following property regards N_u the number of exceeding over a threshold *u* quite high.

Proposition 1 : *Number of exceeding over a high threshold*

The number of exceeding over the threshold u_n *in a sample of length* n *is asymptotically distributed as a Poisson distribution as long as the probability to excess* u_n *decreases proportionally to the inverse of the length of the sample. Formally, we have*

$$\lim_{n \to \infty} n\bar{F}(u_n) = \tau \Rightarrow N_{u_n} \to_d \mathcal{P}(\tau).$$

Indeed, since $\Pr\left[N_{u_n} = k \right] = \Pr\left[\sum_{i=1}^{n} 1_{X_i > u_n} = k \right]$ and that the $X_1, X_2, \ldots$ are independent and identically distributed, we have:

$$\Pr\left[N_{u_n}=k\right]=C_n^k\overline{F}^k\left(u_n\right)\left(1-\overline{F}\left(u_n\right)\right)^{n-k}, \text{ when } n\geq k\ .$$

This equation can be re-written as so:

$$\begin{aligned}\Pr\left[N_{u_n}=k\right]&=C_n^k\overline{F}^k\left(u_n\right)\left(1-\overline{F}\left(u_n\right)\right)^{n-k}\\&=\frac{1}{\left(1-\overline{F}\left(u_n\right)\right)^k}\left(1-\frac{1}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)\overline{F}^k\left(u_n\right)\left(1-\overline{F}\left(u_n\right)\right)^n.\end{aligned}$$

When $n\to\infty$, $n\overline{F}\left(u_n\right)\to\tau$, $\overline{F}\left(u_n\right)\to 0$ and $j/n\to 0$ when $j\in\left\{0,\ldots,k-1\right\}$, then

$$\lim_{n\to\infty}\Pr\left[N_{u_n}=k\right]=\frac{\tau^k}{k!}\lim_{n\to\infty}\left(1-\frac{n\overline{F}\left(u_n\right)}{n}\right)^n=\frac{\tau^k}{k!}e^{-\tau},$$

the last equation can be demonstrated with the Taylor expansion $\ln\left(1-n\overline{F}\left(u_n\right)/n\right)^n$.

Appendix C : Estimation of the tail parameter

The thickness of the tail of the distribution function *F* is defined with the parameter ξ of the MAD distribution from which it belongs. Hence, the distributions that belong to the MAD of the Fréchet's distribution $\left(\xi>0\right)$ have their tail distribution that decreases according to the power whereas the distributions of the Gumbel's MAD $\left(\xi=0\right)$ have their tail distribution exponentially decreasing.

Since the estimators of the extreme percentiles use this tail parameter f_i, its estimation must be as precise as possible.

This section aims at exposing the main methods to estimate the thickness of the tail distribution. The estimators of Pickands (1975), Hill (1975) and Dekkers and al. (1989) are hence presented.

C.1 Parametric methods

C.1.1 Adjustment to the maximum distribution

The Fisher-Tippett-Gnedenko theorem provides the asymptotical distribution of $X_{1,n}$. If we had realizations of the r.v., that is of *m* samples allowing to observe *m* realizations $x_{1,n}^{(1)},\ldots,x_{1,n}^{(m)}$ of $X_{1,n}$, the maximum likelihood method enables to estimate the parameters of the limit distribution, and in particular ξ.

C.1.2 Adjustment of the limit distribution of the excesses

Let us consider a distribution *F* belonging to the maximum attraction domain of G_ξ. According to the Pickands-Balkema-de Haari theorem, a positive function $\beta(.)$ exists such as

$$\lim_{u \to x} \sup_{x>0} \left\{ \left| {}_u F(x) - H_{\xi,\beta(u)}(x) \right| \right\} = 0 .$$

In particular, the parameter ξ of the generalised Pareto distribution $H_{\xi,\beta}$ is the same as the one of G_ξ. Thus, for a high threshold *u*, the distribution $H_{\xi,\beta(u)}$ is a good approximation of F_u from which we have N_u observations available $X_{N_u,n} - u, \ldots, X_{1,n} - u$. The techniques presented in the paragraph A.3 allow us to estimate ξ.

Particularly, if $\xi > -1/2$ is estimated by the maximum likelihood method, Smith (1987) shows that the asymptotical variance of this estimator is $(1+\xi)^2 / N_u$.

C.2 Half parametric methods

C.2.1 Pickands estimator

When $k/n \to 0$, the Pickands estimator, defined with

$$\xi_k^P = \frac{1}{\ln 2} \ln \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4K,n}} ,$$

is a converging estimator of ξ. Moreover, under some supplementary conditions on *k* and *F* (see Dekkers and de Haan (1989)), it is asymptotically Gaussian:

$$\sqrt{k}\left(\xi_k^P - \xi\right) \to_d \mathcal{N}\left(0, \frac{\xi^2\left(2^{\xi+1}+1\right)}{\left(2\left(2^\xi - 1\right)\ln 2\right)^2}\right).$$

C.2.2 Hill estimator

When $k/n \to 0$, the Hill estimator, defined with

$$\xi_k^H = \frac{1}{k} \sum_{j=1}^{k} \ln \frac{X_{j,n}}{X_{k,n}} ,$$

is a converging estimator of ξ. Moreover, under some conditions on *k* and *F* (see de Haan and Peng (1998)), it is asymptotically Gaussian:

$$\sqrt{k}\left(\xi_k^H - \xi\right) \to_d \mathcal{N}\left(0, \xi^2\right).$$

More efficient than the Pickands estimator (see the report on asymptotical variances), the Hill estimator is however useful only with the Fréchet's distributions $(\xi > 0)$.

C.2.3 Dekkers-Einmahl-de Haan estimator

The Dekkers-Einmahl-de Haan estimator is an extent of the Hill estimator to all the extreme distributions $(\xi \in \mathbb{R})$.

When $k/n \to 0$, the Dekkers-Einmahl-de Haan estimator, defined with

$$\xi_k^{DEdH} = 1 + \xi_k^{(1)} - \frac{1}{2}\left(1 - \frac{\left(\xi_k^{(1)}\right)^2}{\xi_k^{(2)}}\right)^{-1},$$

where $\xi_k^{(i)} = \frac{1}{k}\sum_{j=1}^{k}\left(\ln \frac{X_{j,n}}{X_{k+1,n}}\right)^i$, is a converging estimator of ξ. Moreover, under some conditions on k and F (see Dekkers and al. (1989)), it is asymptotically Gaussian:

$$\sqrt{k}\left(\xi_k^{DEdH} - \xi\right) \to_d \mathcal{N}\left(0, 1 + \xi^2\right).$$

This estimator is also known as the estimator of moments, since the $\xi_k^{(i)}$ can be interpreted as the empirical moments.

C.2.4 Number of observations needed

The results on the estimator of the tail index previously exposed are asymptotic: they are found when $k \to \infty$ and $k/n \to 0$. As in practice, only a finite number *n* of observations are available, we have to choose *k* so as to have a tour disposal enough statistical tools (the $X_{k,n}, \ldots, X_{1,n}$) and still stay in the tail distribution $(k \ll n)$.

In particular, the Hill estimator satisfies the asymptotical property

$$\sqrt{k}\left(\xi_k^{H} - \xi\right) \to_d \mathcal{N}\left(0, \xi^2\right),$$

when $k \to \infty$ with a *certain growing rate* according to *n*, we could be tempted to choose *k* as high as possible so as to minimize the mean square error engendered by ξ_k^H. However, the behaviour of the second order of the slow variation function $\mathcal{L}_F$ introduced in the Tauberian theorem engenders a bias when *k* is too high. Solutions to fix *k* so as to have a non-biased asymptotical estimator were suggested by Goldie and Smith (1987) and then by de Haan and Peng (1998).

About the Hill estimator for functions in the Fréchet's MAD, de Haan and Peng (1998) have suggested to use the number of observations k^* that reduces the mean square error of the Hill estimator, that is

$$k^*(n) = \begin{cases} 1 + n^{2\xi/(2\xi+1)}\left(\dfrac{(1+\xi)^2}{2\xi}\right)^{1/(2\xi+1)}, & \text{si } \xi \in]0;1[\\ 2n^{2/3}, & \text{si } \xi > 1. \end{cases}$$

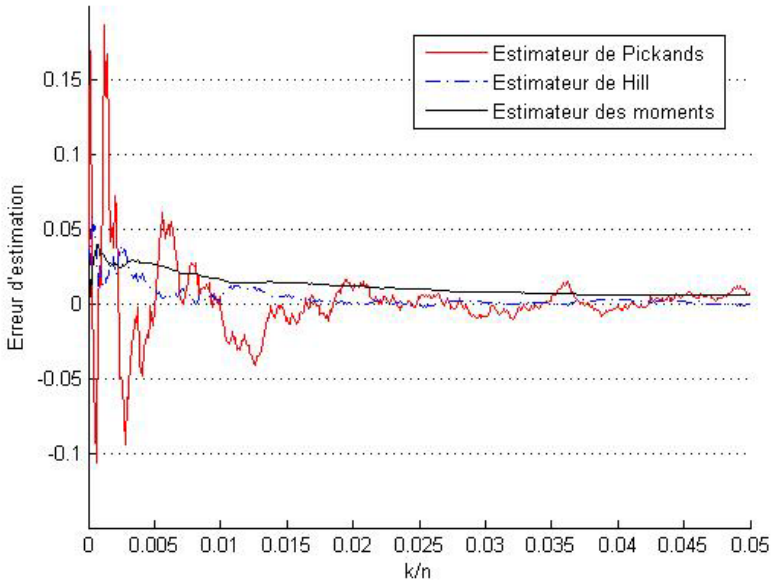
However, we notice that k^* depends of ξ which is not observed. Moreover, this criterion has to do only with the Hill estimator.

In practice, it is advisable to follow the process suggested by Embrechts and al. (1997) which consists in drawing the estimators according to k/n and to choose k in an interval where the estimators' line is almost horizontal (see the Figure 9 detailed further).

C.2.5 Illustration

The Figure 9 illustrates the different estimators of the tail parameters of a first type Pareto distribution (with a distribution function of $F(x) = 1 - (x_0/x)^{\alpha}$, when $x > x_0$). This distribution family belongs to the Fréchet's MAD, the Pickands', Hill's and de Dekkers-Einmahl-de Haan's estimators have been drawn according to k.

Figure 9 – Estimate of the thickness of the tail of a Pareto distribution



The supremacy of the Hill estimator over the Pickands' and de Dekkers-Einmahl-de Haan's estimators is observed. Furthermore, when $k < 0,02 \times n$, the Hill estimator is relatively volatile. We would then have to use 2.5% of the extreme data to estimate the thickness of the tail distribution.

Bibliography

Blum K. A., Otto D. J. (1998) « Best estimate loss reserving : an actuarial perspective », *CAS Forum Fall* 1, 55-101.

Bottard S. (1996) « Application de la méthode du Bootstrap pour l'estimation des valeurs extrêmes dans les distributions de l'intensité des séismes », *Revue de statistique appliquée* 44 (4), 5-17.

Christofiersen P., Hahn J., Inoue, A. (2001) « Testing and comparing value-at-risk measures », *Journal of Empirical Finance* 8 (3), 325-42.

Coles S., Powell E. (1996) « Bayesian methods in extreme value modelling : a review and new developments », *Internat. Statist. Rev.* 64, 119-36.

Davidson R., MacKinnon J.G. (2004) « Bootstrap Methods in Econometrics », working paper.

de Haan L., Peng L. (1998) « Comparison of tail index estimators », *Statistica Neerlandica* 52 (1), 60-70.

Deelstra G., Janssen J. (1998) « Interaction between asset liability management and risk theory », *Applied Stochastic Models and Data Analysis* 14, 295-307.

Dekkers A., de Haan L. (1989) « On the estimation of the extreme-value index and large percentile estimation », *Annals of Statistics* 17, 1795-832.

Dekkers A., Einmahl J., de Haan L. (1989) « A moment estimator for the index of an extreme-value distribution », *Annals of Statistics* 17, 1833-55.

Denuit M., Charpentier A. (2005) *Mathématiques de l'assurance non-vie. Tome 2 : tarification et provisionnement*, Paris : Economica.

Diebolt J., El-Aroui M., Garrido S., Girard S. (2005a) « Quasi-conjugate bayes estimates for gpd parameters and application to heavy tails modelling », *Extremes* 8, 57-78.

Diebolt J., Guillou A., Rached I. (2005b) « Approximation of the distribution of excesses using a generalized probability weighted moment method », *C. R. Acad. Sci. Paris*, Ser. I 340 (5), 383-8.

Efron B. (1979) « Bootstrap methods: Another look at the Jackknife », *Ann. Statist*., 71-26.

Efron B. (1987) « Better bootstrap confidence intervals », *Journal of the American Statistical Association* 82, 171-200.

Efron B., Tibshirani R. J. (1993) *An introduction to the bootstrap*, Chapman & Hall.

Embrechts P., Klüppelberg C., Mikosch T. (1997) *Modelling Extremal Events for Insurance and Finance*, Berlin: Springer Verlag.

England P.D., Verall R.J. (1999) « Analytic and bootstrap estimates of prediction errors in claim reserving », *Insurance: mathematics and economics* 25, 281-93.

Fedor M., Morel J. (2006) « Value-at-risk en assurance : recherche d'une méthodologie à long terme », *Actes du 28e congrès international des actuaires*, Paris.

Goldie C., Smith R. (1987) « Slow variation with remainder: a survey of the theory and its applications », *Quarterly Journal of Mathematics Oxford* 38 (2), 45-71.

Hill B. (1975) « A simple general approach to inference about the tail of a distribution », *Annals of Statistics* 3, 1163-74.

Horowitz J.L. (1997) « Bootstrap methods in econometrics: theory and numerical performance », in *Advances in Economics and Econometrics:*

Theory and Application, volume 3, 188-222, D. M. Kreps, K. F. Wallis (eds), Cambridge: Cambridge University Press.

Hosking J. R., Wallis J. R. (1987) « Parameter and percentile estimation for the generalized pareto distribution », *Technometrics* 29, 339-49.

Jal P., Partrat Ch. (2004) « Evaluation stochastique de la provision pour sinistres », Conférence scientifique de l'Institut des Actuaires, Paris, 20 janvier 2004.

Merton R. (1976) « Option pricing when underlying stock returns are discontinuous », *Journal of Financial Economics* 3, 125-44.

Partrat C., Besson J.L. (2005) *Assurance non-vie. Modélisation, simulation*, Paris : Economica.

Pickands J. (1975) « Statistical inference using extreme orders statistics », *Annals of Statistics* 3, 119-31.

Planchet F., Thérond P.E. (2005b) « L'impact de la prise en compte des sauts boursiers dans les problématiques d'assurance », *Proceedings of the 15th AFIR Colloquium*, Zürich.

Quenouille M. (1949) « Approximate tests of correlation in time series », *Journal of the Royal Statistical Society*, Soc. Series B, 11, 18-84.

Smith R. L. (1987) « Estimating tails of probability distributions », *Annals of Statistics* 15, 1174-207.

Thérond P.E., Planchet F. (2007) « Provisions techniques et capital de solvabilité d'une compagnie d'assurance : méthodologie d'utilisation de Value-at-Risk », *Assurances et gestion des risques* 74 (4), 533-63.

Windcliff H. Boyle, P. (2004) « The 1/n pension investment puzzle », *North American Actuarial Journal* 8 (3), 32-45.

Zajdenweber D. (2000) *Économie des extrêmes*, Paris : Flammarion.

Zelterman D. (1993) « A semiparametric Bootstrap technique for simulating extreme order statistics », *Journal of American Statistical Association* 88 (422), 477-85.