

ATELIER DE THÉORIE DES VALEURS EXTRÊMES

APPLICATION EN FINANCE QUANTITATIVE

Pierre Clauss

Ensaï
Filière Génie Statistique

TABLE DES MATIÈRES

INTRODUCTION	3
1 THÉORÈMES DE LA TVE	4
1.1 Théorème de Fisher-Tippet	4
1.2 Théorème de Balkema-de Haan-Pickands	5
2 MÉTHODES STATISTIQUES ASSOCIÉES À LA TVE	7
2.1 Modélisation paramétrique de la distribution des maxima par blocs	7
2.1.1 Sélection de la taille des blocs	7
2.1.2 Estimation du modèle BM	7
2.2 Modélisation paramétrique de la distribution des excès	8
2.2.1 Sélection du seuil	8
2.2.2 Estimation du modèle POT	9
3 APPLICATION EN FINANCE QUANTITATIVE	11
Préambule : la notion de mesure du risque	11
3.1 Définition de la Value-at-Risk	12
3.2 VaR classiques	12
3.2.1 VaR historique	12
3.2.2 VaR paramétrique gaussienne	13
3.3 VaR TVE	13
3.3.1 Méthodologie du calcul d'une VaR TVE	13
3.3.2 Modèle dynamique	15
3.3.3 Méthodologie du calcul d'une VaR TVE dynamique	16
CONCLUSION	17
BIBLIOGRAPHIE	18

INTRODUCTION

Cet atelier de 5h faisant suite au cours de Théorie des Valeurs Extrêmes (TVE) est composé de 3 parties :

1. un rappel des théorèmes probabilistes essentiels à la compréhension de la TVE : les théorèmes de Fisher-Tippet et de Balkema-de Haan-Pickands,
2. les méthodes statistiques associées : les méthodes Block Maxima et Peaks Over Threshold,
3. et enfin une application à la mesure du risque de marché supporté par un portefeuille d'investissement. Cette application est bien entendu généralisable à d'autres domaines.

La *philosophie* de cet atelier est d'aborder en profondeur les techniques statistiques de la TVE et de les appliquer à des données. Pour cela, une partie importante de l'atelier est dédiée à la programmation informatique à l'aide essentiellement du logiciel R¹, au travail sur les données, ainsi qu'à l'étude d'un article scientifique.

1. R Development Core Team (2008). R : A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

CHAPITRE 1

THÉORÈMES DE LA TVE

La Théorie des Valeurs Extrêmes s'intéresse non pas à la modélisation totale d'une distribution mais seulement aux queues de cette distribution à partir de lois spécifiques propres que nous allons préciser dans ce qui suit (cf. Embrechts, Kluppelberg, Mikosch [6] et Beirlant, Goegebeur, Segers, Teugels [2] pour plus d'approfondissements).

Deux théorèmes sont essentiels à la compréhension de la Théorie des Valeurs Extrêmes : celui de Fisher-Tippet [8] et celui de Balkema-de Haan-Pickands [1, 12].

1.1 Théorème de Fisher-Tippet

Définition 1.1 Nous supposons avoir une suite initiale d'observations Z_1, \dots, Z_n issue d'une fonction de distribution inconnue F (aucune hypothèse n'est présupposée). Cet échantillon peut être séparé en k blocs disjoints de même longueur s (un bloc peut correspondre à un mois, un an, etc.). Les données $Z_1^{(i)}, \dots, Z_s^{(i)}$, $i = 1, \dots, k$ sont iid avec comme fonction de distribution F . Nous nous intéressons aux maxima de ces k blocs $Y_i^s = \max(Z_1^{(i)}, \dots, Z_s^{(i)})$, qui forment l'échantillon de données supposées iid Y_1^s, \dots, Y_k^s .

La loi fondamentale à la modélisation des maxima est la Generalized Extreme Value (GEV) définie par la fonction de répartition suivante :

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

où x est tel que $1 + \xi x > 0$. ξ est le paramètre de forme ou de queue. La GEV rassemble trois distributions particulières : si $\xi > 0$, c'est la distribution de Fréchet ; si $\xi < 0$, nous avons la distribution de Weibull ; enfin, $\xi = 0$ donne la distribution de Gumbel. La famille peut être étendue en ajoutant les paramètres de localisation μ et de dispersion $\sigma > 0$: la GEV $H_{\xi, \mu, \sigma}(x)$ est en fait $H_\xi((x - \mu)/\sigma)$.

Le théorème essentiel à la modélisation des maxima est celui de Fisher-Tippet [8] prouvé en 1928.

Théorème 1.1 Supposons que nous pouvons trouver deux suites d'entiers réels $a_i > 0$ et b_i telles que :

$$a_i^{-1}(Y_i^s - b_i) \xrightarrow{L} H \text{ pour } s \rightarrow \infty$$

ce qui est équivalent à

$$\mathbb{P}\{a_i^{-1}(Y_i^s - b_i) \leq y\} = F^s(a_i y + b_i) \rightarrow H(y) \text{ pour } s \rightarrow \infty \quad (1.1)$$

avec H une loi non dégénérée. $a_i^{-1}(Y_i^s - b_i)$ sont les k maxima normalisés. Alors, F est dans le maximum domain of attraction (MDA) de H , ce qui s'écrit $F \in \text{MDA}(H)$.

Fisher-Tippett montre alors que $F \in \text{MDA}(H)$ si, et seulement si, H est du type de H_ξ . La GEV est donc la seule distribution limite non dégénérée pour un échantillon de maxima normalisés ¹.

1.2 Théorème de Balkema-de Haan-Pickands

Définition 1.2 Nous supposons toujours avoir une suite d'observations Z_1, \dots, Z_n , issue d'une fonction de distribution inconnue F , cette fois-ci présumées iid. Soit z_0 la limite supérieure finie ou infinie du support de la loi F . Nous nous intéressons alors au nombre N_u d'extrêmes (strictement positifs) qui excèdent un seuil élevé u , et plus précisément aux excès au-dessus du seuil u , $X_i = Z_i - u$, qui forment l'échantillon des excès X_1, \dots, X_{N_u} , supposés iid. Nous définissons la fonction de distribution des excès au dessus du seuil u par

$$F_u(x) = \mathbb{P}\{Z - u \leq x | Z > u\} = \mathbb{P}\{X \leq x | Z > u\} = \frac{F(x + u) - F(u)}{1 - F(u)} \quad (1.2)$$

pour $0 \leq x < z_0 - u$. $F_u(x)$ est donc la probabilité qu'un extrême excède le seuil u par une quantité inférieure ou égale à x , sachant que le seuil est dépassé.

La loi essentielle à la modélisation des excès est la Generalized Pareto Distribution (GPD) définie par la fonction de répartition suivante :

$$G_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - \exp(-x) & \text{si } \xi = 0 \end{cases}$$

où le support est $x \geq 0$ lorsque $\xi \geq 0$ et il devient $0 \leq x \leq -1/\xi$ lorsque $\xi < 0$. La GPD regroupe trois distributions selon les valeurs du paramètre de forme. Lorsque $\xi > 0$, c'est la loi Pareto usuelle ; lorsque $\xi < 0$, nous avons la loi de Pareto de type II ; et $\xi = 0$ donne la loi exponentielle. La famille peut être étendue en ajoutant le paramètre de localisation μ : la GPD $G_{\xi, \mu, \sigma}(x)$ est en fait $G_\xi((x - \mu)/\sigma)$. Si la loi des excès est $G_{\xi, \sigma}(x)$, alors celle des extrêmes est $G_{\xi, \sigma}(x - u)$.

Le théorème essentiel à la modélisation des excès est celui de Balkema-de Haan-Pickands [1, 12] déterminé en 1974-1975.

Théorème 1.2 Pour une certaine classe de distributions², la GPD est la distribution limite de la distribution des excès lorsque le seuil tend vers z_0 . Formellement, nous pouvons trouver une fonction positive mesurable $\sigma(u)$ telle que

$$\lim_{u \rightarrow z_0} \sup_{0 \leq x < z_0 - u} |F_u(x) - G_{\xi, \sigma(u)}(x)| = 0 \quad (1.3)$$

si, et seulement si, $F \in \text{MDA}(H_\xi)$.

En résumé, nous avons :

Si pour une distribution F inconnue³, l'échantillon des maxima normalisés converge en loi vers une distribution non dégénérée, alors il est équivalent de dire que F est dans le *maximum domain of attraction* de la GEV H_ξ (**Fisher-Tippett**). Dans ce cas, il s'en déduit que la distribution des excès au-dessus d'un seuil élevé converge vers une GPD $G_{\xi, \sigma}$ lorsque le seuil tend vers la limite supérieure du support de F (**Balkema-de Haan-Pickands**).

1. Par exemple, la distribution de Student est dans le *maximum domain of attraction* de la distribution de Fréchet qui présente des queues épaisses. La distribution Gaussienne est quant à elle dans le MDA de la loi de Gumbel qui présente des queues de taille moyenne.

2. Cette classe est large : elle regroupe toutes les distributions continues classiques.

3. Si F était connue, nous n'aurions pas besoin de nous intéresser au comportement asymptotique des maxima puisque F^n serait très simplement identifiable.

A partir de ces deux résultats de probabilité, nous obtenons deux méthodes différentes de modélisation statistique : le premier théorème étudie le comportement asymptotique d'un échantillon de maxima (modélisation *Block Maxima*) ; alors que le second étudie celui d'un échantillon d'excès au-dessus d'un seuil élevé (modélisation *Peaks Over Threshold*).

CHAPITRE 2

MÉTHODES STATISTIQUES ASSOCIÉES À LA TVE

Deux méthodes principales de modélisation des événements rares sont possibles : la méthode *Block Maxima* (BM) qui modélise la distribution des extrêmes par la distribution *Generalized Extreme Value* (GEV) ; et la méthode *Peaks Over Threshold* (POT) qui modélise la distribution des excès au-dessus d'un seuil¹ (en anglais *threshold*) élevé par la *Generalized Pareto Distribution* (GPD).

2.1 Modélisation paramétrique de la distribution des maxima par blocs

Cette modélisation des queues de distribution s'appuie sur le théorème de Fisher-Tippet et nous supposons que l'échantillon de maxima suit exactement une loi GEV.

2.1.1 Sélection de la taille des blocs

Il n'y a pas d'outils statistiques d'aide à la sélection de la taille des blocs. Il faut cependant que la condition asymptotique (1.1) de la section 1.1 soit vérifiée et donc que s soit suffisamment grand. Mais, il faut aussi que nous ayons un nombre suffisant de maxima pour que l'estimation des paramètres de la GEV soit assez précise. Il est donc courant de prendre $s = 21$ (pour un mois) ou $s = 251$ (pour un an), valeurs qui représentent un bon compromis.

2.1.2 Estimation du modèle BM

A partir de l'échantillon des maxima construit précédemment, nous pouvons estimer les paramètres de la GEV. Deux méthodes d'estimation sont envisageables : l'Estimation par Maximum de Vraisemblance (EMV) et celle par les Moments Pondérés (EMP). Ici, nous ne verrons que l'EMV (nous étudierons l'EMP dans le cadre de la méthode POT, cf. section suivante).

Soit l'échantillon de maxima supposé iid $Y = (Y_1, \dots, Y_k)$ et $h_{\xi, \mu, \sigma}$ la densité de la loi GEV $H_{\xi, \mu, \sigma}$. Cette dernière s'écrit pour $\xi \neq 0$:

$$h_{\xi, \mu, \sigma}(y) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{-\frac{1+\xi}{\xi}} \exp \left\{ - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}$$

1. Dans la littérature, il se peut que la méthode POT comprenne aussi la modélisation de la fréquence des excès : ceci est une extension du modèle POT classique, qui décrit le comportement des excès ainsi que leur fréquence par un *point process* de Poisson à deux dimensions. Nous ne l'étudierons pas.

Et la vraisemblance de l'échantillon Y est égale à :

$$\mathcal{L}(\xi, \mu, \sigma; Y) = \prod_{i=1}^n h_{\xi, \mu, \sigma}(Y_i) \text{ (car l'échantillon est iid)}$$

Il est fait appel à des procédures numériques (algorithme de Quasi-Newton) pour la maximisation de la vraisemblance. Alors le calcul des estimateurs ne pose pas de sérieux problèmes. En revanche, rien ne nous assure de leurs régularités (estimateurs asymptotiquement efficaces et normaux) surtout lorsque l'échantillon est de petite taille. Smith [15] montre qu'il suffit que $\xi \geq -0.5$ ² pour que les conditions de régularité de l'EMV soient remplies.

Dans le cas où $\xi = 0$, la log-vraisemblance est égale à :

$$l(0, \mu, \sigma; Y) = -n \ln \sigma - \sum_{i=1}^n \exp\left(-\frac{Y_i - \mu}{\sigma}\right) - \sum_{i=1}^n \frac{Y_i - \mu}{\sigma}$$

En dérivant cette fonction relativement aux deux paramètres, nous obtenons le système d'équations à résoudre suivant :

$$\begin{aligned} n - \sum_{i=1}^n \exp\left(-\frac{Y_i - \mu}{\sigma}\right) &= 0 \\ n + \sum_{i=1}^n \frac{Y_i - \mu}{\sigma} \left[\exp\left(-\frac{Y_i - \mu}{\sigma}\right) - 1 \right] &= 0 \end{aligned}$$

Précisons cependant qu'il n'existe pas de solution explicite à ces équations de maximisation (utilisation de méthodes numériques, type algorithmes de Newton-Raphson).

2.2 Modélisation paramétrique de la distribution des excès

Cette modélisation des queues de distribution s'appuie sur le théorème de Balkema-de Haan-Pickands et nous supposons que l'échantillon des excès au-dessus du seuil u suit exactement une loi GPD.

2.2.1 Sélection du seuil

Avant de pouvoir estimer le modèle, il faut construire un échantillon adéquat. Il nous faut trouver un seuil u de sélection des données extrêmes suffisamment élevé pour que l'approximation asymptotique (1.3) de la section 1.2 soit applicable (si nous choisissons un seuil trop bas, les estimations seront biaisées) et au-dessus duquel nous conservons assez de données pour des estimations précises (si le seuil est trop élevé, les écarts-types des estimateurs seront très importants).

Un des outils de choix du seuil est le graphe de la *sample mean excess function* $e_n(u)$ (ME-plot).

Définition 2.1 *Le ME-plot est défini de la manière suivante :*

$$\{(u, e_n(u)), Z_{n:n} < u < Z_{1:n}\}$$

où $Z_{1:n}$ et $Z_{n:n}$ sont respectivement les maximum et minimum de l'échantillon. Et $e_n(u)$ est définie par

$$e_n(u) = \frac{\sum_{i=1}^n (Z_i - u)^+}{\sum_{i=1}^n 1_{\{Z_i > u\}}} = \frac{1}{N_u} \sum_{i=1}^n (Z_i - u)^+$$

2. Ce qui est le cas généralement des distributions des rentabilités des actifs financiers.

c'est-à-dire la somme des excès au-dessus du seuil u divisé par le nombre N_u de données qui excèdent u . La sample mean excess function $e_n(u)$ est l'estimateur empirique de la mean excess function

$$e(u) = \mathbb{E}[Z - u | Z > u]$$

Comment interprète-t-on ce graphe ? Il faut savoir tout d'abord que la *mean excess function* de la GPD est

$$e(u) = \frac{\sigma + \xi u}{1 - \xi} \quad (2.1)$$

où $\sigma + \xi u > 0$. Ainsi, si le ME-plot semble avoir un comportement linéaire au-dessus d'une certaine valeur de u , cela signifie que les excès au-dessus de ce seuil suivent une GPD.

Dans l'exemple de la Figure 2.1 (pertes en valeur absolue d'un indice boursier), nous pouvons définir le seuil à 1 : en effet, cette valeur correspond à un judicieux arbitrage entre le souhait d'avoir un seuil élevé et celui d'obtenir un échantillon d'excès de taille importante.

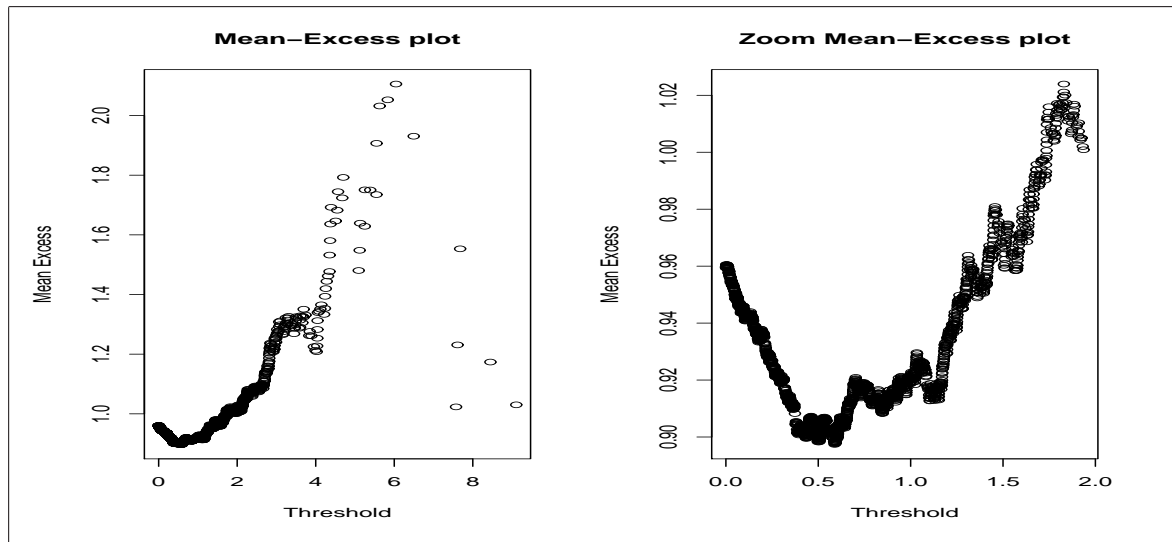


FIGURE 2.1 – Mean-Excess plots pour la queue des pertes

2.2.2 Estimation du modèle POT

Deux méthodes d'estimation sont ici encore envisageables : l'Estimation par Maximum de Vraisemblance (EMV) et celle par les Moments Pondérés (EMP).

Les estimateurs du Maximum de Vraisemblance

Supposons que notre échantillon des excès $X = (X_1, \dots, X_{N_u})$ est iid avec comme fonction de distribution la GPD G . La fonction de densité g de G est alors pour $\xi \neq 0$

$$g(x) = \frac{1}{\sigma} \left(1 + \xi \frac{x}{\sigma}\right)^{-\frac{1}{\xi} - 1}$$

La log-vraisemblance est donc égale à

$$l(\xi, \sigma; X) = -N_u \ln \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{N_u} \ln \left(1 + \frac{\xi}{\sigma} X_i\right)$$

En dérivant cette fonction en ξ et σ , nous obtenons les équations de maximisation à partir desquelles nous calculons les estimateurs du Maximum de Vraisemblance $(\hat{\xi}_{N_u}, \hat{\sigma}_{N_u})$ (à l'aide de méthodes numériques). Et pour $\xi = 0$, nous avons

$$g(x) = \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right)$$

$$l(0, \sigma; X) = -N_u \ln \sigma - \frac{1}{\sigma} \sum_{i=1}^{N_u} X_i$$

Nous obtenons alors $\hat{\sigma}_{N_u} = \sum_{i=1}^{N_u} X_i / N_u$ qui n'est autre que la moyenne empirique des excès (pour $\xi = 0$, la GPD est la loi exponentielle).

Les estimateurs des Moments Pondérés

Il peut arriver que certains moments n'existent pas, ne soient pas finis. Au lieu de la Méthode des Moments, nous utilisons alors la Méthode des Moments Pondérés. Définissons, avec r l'ordre du moment

$$w_r(\xi, \sigma) = \mathbb{E}[X \bar{G}_{\xi, \sigma}^r(X)]$$

où $\bar{G}_{\xi, \sigma} = 1 - G_{\xi, \sigma}$ avec G la loi GPD. Alors

$$w_r(\xi, \sigma) = \int_{-\infty}^{+\infty} x \bar{G}_{\xi, \sigma}^r(x) d\bar{G}_{\xi, \sigma}(x) = \int_0^1 \bar{G}_{\xi, \sigma}^{-1}(y) y^r dy = \int_0^1 \frac{\sigma}{\xi} (y^{-\xi} - 1) y^r dy$$

Nous obtenons grâce à la dernière formulation et après quelques calculs

$$w_r(\xi, \sigma) = \frac{\sigma}{(r+1)(r+1-\xi)}$$

ce qui nous permet d'obtenir avec $r = 0$ et $r = 1$

$$\sigma = \frac{2w_0 w_1}{w_0 - 2w_1}$$

$$\xi = 2 - \frac{w_0}{w_0 - 2w_1}$$

Nous avons aussi

$$\hat{w}_r(\xi, \sigma) = \frac{1}{n} \sum_{i=1}^{N_u} X_i \hat{F}^r(X_i), \quad r = 0, 1$$

où \hat{F} est la fonction de répartition empirique de l'échantillon X_1, \dots, X_{N_u} . Pour estimer ξ et σ , nous remplaçons w_r par \hat{w}_r pour $r = 0, 1$.

Hosking et Wallis [11] ont montré que lorsque $0 \leq \xi \leq 0.4$ et pour des échantillons de taille petite, l'EMP obtient des estimateurs plus précis que l'EMV (avec des écarts-types plus faibles). Néanmoins, cette différence s'atténue avec l'augmentation de la taille de l'échantillon. En outre, Rootzén et Tajvidi [13] révèlent que pour $\xi \geq 0.5$, l'EMP calcule des estimateurs fortement biaisés contrairement aux estimateurs de l'EMV qui sont efficaces. Enfin, pour $\xi \geq -0.5$, les conditions de régularité de l'EMV sont remplies et les estimateurs du Maximum de Vraisemblance $(\hat{\xi}_{N_u}, \hat{\sigma}_{N_u})$ calculés sur l'échantillon des N_u excès sont asymptotiquement normaux (Hosking et Wallis).

CHAPITRE 3

APPLICATION EN FINANCE QUANTITATIVE

Préambule : la notion de mesure du risque

Le risque de marché correspond à un risque de perte des portefeuilles des investisseurs dues aux variations des marchés financiers : marchés des instruments de base (actions, obligations, devises, matières premières) mais aussi marchés des produits dérivés (contrats à terme, options). Pour gérer le risque de marché, il faut donc mesurer de manière précise ce risque pour mieux l'appréhender, puis développer des outils de réduction et de couverture des risques.

Une première mesure du risque est la **sensibilité** du produit financier détenu relativement à son marché. Ainsi, puisque le risque de marché est le risque de perte consécutivement aux évolutions des marchés, si, naïvement, nous possédons une action européenne et que le marché des actions européen subit une baisse, il semble naturel de penser qu'il va en être de même pour notre action. Cette première mesure s'assimile à la sensibilité des portefeuilles ou titres détenus relativement à des facteurs de risque de marché. Ainsi, des modèles de risque se sont développés comme le Capital Asset Pricing Model (cf. Sharpe [14]) pour le plus connu. Ils mesurent les sensibilités (coefficients de régression) des titres ou portefeuilles de titres financiers à des facteurs de risque tels le marché dans son ensemble par exemple.

Néanmoins, le *risk manager* a besoin d'une mesure plus synthétique du risque encouru par son portefeuille d'investissement. En effet, lorsque plusieurs instruments très différents composent le portefeuille, il est difficile d'agrèger les nombreuses sensibilités. Nous pouvons ainsi appréhender le risque à partir des profils de la dispersion des pertes et profits des actifs. Nous pouvons faire ressortir deux mesures de risque à partir de la distribution des rentabilités des actifs : la **volatilité** et la **Value-at-Risk** ou **VaR**.

Ces mesures ne captent pas la même chose :

- la volatilité mesure la dispersion moyenne des rentabilités autour de leur moyenne : en effet, son expression en fonction du vecteur des rentabilités R est $\sigma = \sqrt{\mathbb{E}[(R - \mathbb{E}(R))^2]}$. Elle prend donc en compte toutes les rentabilités, positives ou non, extrêmes ou non.
- la Value-at-Risk est le quantile de perte déterminant la plus grande perte que peut subir un portefeuille avec une probabilité d'occurrence faible sur un horizon donné : elle mesure ainsi le risque extrême.

La logique entre les deux mesures est donc différente. En effet, la volatilité peut être élevée et seulement capturer des risques **moyens**, certes importants, mais pas **extrêmes**. Tout l'enjeu d'une mesure du risque synthétique pertinente est d'estimer correctement la perte potentielle que peut subir un portefeuille. Or un actif peut avoir une volatilité de 20% mais ne pas expérimenter de pertes très fortes, avec un maximum par exemple de 15% de perte sur une journée. Au contraire, un actif avec une volatilité de 15% peut connaître des pertes, certes rares, de plus de 30%. Le second actif nous semble donc plus risqué, même si ses risques moyens s'avèrent être moins importants que ceux du premier.

En fait, déterminer le risque par le moment d'ordre 2 (la volatilité) présuppose que les moments suivants (skewness et kurtosis) ne nécessitent pas d'être incorporés dans une mesure de risque. Le paradigme sous-jacent est la normalité des rentabilités d'un actif financier (la loi Normale étant caractérisée par les deux premiers moments). Utiliser la Value-at-Risk permet de résoudre ce problème car un quantile de distribution n'est pas une mesure moyenne de la dispersion mais prend seulement en compte les pertes qui nous intéressent ici, c'est-à-dire les pertes extrêmes.

Nous venons d'introduire le fait que la volatilité n'est certainement pas la meilleure mesure de risque extrême. C'est pour cela que la Value-at-Risk est utilisée par la plupart de risk managers. Nous allons dans cette partie étudier plus en détail cette mesure de risque en utilisant les résultats de la TVE.

3.1 Définition de la Value-at-Risk

La VaR représente une perte potentielle dont il faut définir la **probabilité** α d'occurrence, appelée aussi niveau de confiance, ainsi que l'**horizon** h de durée d'investissement. Par exemple, une probabilité de 1% pour un horizon h de un jour revient à accepter que la perte potentielle ne dépasse la VaR que 2 à 3 fois par an (1 an = 251 jours ouvrés). L'horizon représente la durée sur laquelle nous estimons le risque, c'est-à-dire le temps nécessaire pour couvrir la position du portefeuille : 1 jour en trading, 1 mois en gestion. Soit V_t^P la valeur en t du portefeuille et V_{t+h}^P la valeur à l'horizon h . La rentabilité du portefeuille $R_{t|t+h}^P$ investi à la date t et détenu sur un horizon de durée h a pour expression $\forall t \geq 1$:

$$R_{t|t+h}^P = \ln \frac{V_{t+h}^P}{V_t^P} \approx \frac{V_{t+h}^P}{V_t^P} - 1.$$

Supposons que $R_{t|t+h}^P$ est issue d'une variable aléatoire notée R^P . La $\text{VaR}(h, \alpha)$ est alors définie par :

$$\mathbb{P}[R^P \leq \text{VaR}(h, \alpha)] = \alpha$$

c'est-à-dire la probabilité que la rentabilité du portefeuille soit inférieure à la VaR sur un horizon h est égale à α ¹.

Définition 3.1 Notons F la fonction de répartition de la variable aléatoire R^P des rentabilités du portefeuille de périodicité h . La Value-at-Risk de probabilité d'occurrence α et d'horizon d'investissement h est alors le quantile défini par :

$$\text{VaR}(h, \alpha) = F^{-1}(\alpha) \tag{3.1}$$

Il nous reste alors à caractériser la forme de F . La première solution consiste à estimer F par sa distribution empirique : nous avons alors la VaR historique. La seconde est d'utiliser un modèle paramétrique : les praticiens utilisent de manière classique la VaR Gaussienne, que nous étudierons. Mais le TVE nous fournira des résultats beaucoup plus précis.

3.2 VaR classiques

3.2.1 VaR historique

Pour la VaR historique, nous estimons F non-paramétriquement par sa distribution empirique \hat{F} . Supposons que nous disposons d'un historique de T observations des rentabilités. La distribution empirique \hat{F} des rentabilités du portefeuille est alors caractérisée à partir du vecteur $(\hat{R}_{1|1+h}^P, \dots, \hat{R}_{T|T+h}^P)'$ et $\hat{F}^{-1}(\alpha)$ correspond au quantile de probabilité α de $(\hat{R}_{1|1+h}^P, \dots, \hat{R}_{T|T+h}^P)'$.

1. Les praticiens peuvent aussi définir le niveau de confiance par $1 - \alpha$ qui correspond de même au quantile de probabilité α : ainsi, une VaR à 99% par exemple correspond au quantile de probabilité $\alpha = 1 - 99\% = 1\%$.

L'avantage de cette méthode est donc de ne pas imposer d'hypothèse sur la loi de distribution des rentabilités à la différence de la méthode précédente.

Mais deux problèmes se posent : tout d'abord la longueur T de l'historique qui, si elle est trop faible, ne fournit pas un calcul précis de la VaR (ceci peut être résolu par des techniques de ré-échantillonnage ou *bootstrap*) ; en outre, l'inconvénient d'une VaR historique est l'*over-fitting* ou encore la dépendance trop forte à l'échantillon sur lequel elle a été déterminée, et donc la difficulté de cette VaR à être robuste dans le temps. C'est pour cela que les VaR paramétriques, certes moins ajustées aux données étudiées, permettent une meilleure prévision.

3.2.2 VaR paramétrique gaussienne

De manière classique, le choix du modèle paramétrique se porte sur la loi Normale. Supposons donc que les rentabilités du portefeuille $R_{t|t+h}^P$ sont issues d'une variable aléatoire gaussienne : ainsi, F est définie par $\mathcal{N}(\mu, \sigma^2)$ avec μ l'espérance des rentabilités et σ^2 la variance des rentabilités.

Nous obtenons ainsi à partir de l'équation (3.1) :

$$\Phi\left(\frac{\text{VaR}(h, \alpha) - \mu}{\sigma}\right) = \alpha$$

avec Φ la fonction de répartition de la loi Normale centrée réduite.

Définition 3.2 La VaR Gaussienne est définie de la manière suivante :

$$\text{VaR}(h, \alpha) = \mu + \sigma\Phi^{-1}(\alpha)$$

Cette méthode repose sur des hypothèses irréalistes pour modéliser les rentabilités des actifs financiers puisque trois faits stylisés caractérisent leur distribution et ne sont pas captés par la modélisation gaussienne :

1. l'asymétrie : l'occurrence de pertes extrêmes est plus forte que celle de profits extrêmes,
2. les queues épaisses.
3. le *clustering* : par exemple, les rentabilités élevées en valeur absolue sont suivies par des rentabilités élevées en valeur absolue,

Les deux premiers faits stylisés peuvent s'observer sur la Figure 3.1 à l'aide des QQ-plot et densité de la queue des pertes de l'indice de marché français CAC 40 relativement à la loi Normale.

3.3 VaR TVE

3.3.1 Méthodologie du calcul d'une VaR TVE

Deux méthodes statistiques de modélisation des queues s'offrent à nous : la méthode BM et la méthode POT. En Finance, néanmoins, une méthode va être préférée à cause du phénomène de *clustering* vu précédemment.

En effet, nous allons privilégier la modélisation des queues de distribution par la méthode *Peaks Over Threshold* avec l'Estimation par Maximum de Vraisemblance qui possède plusieurs avantages sur les autres méthodes. Tout d'abord, elle est assez flexible et réaliste par rapport à la méthode *Block Maxima*, qui ne prend pas en compte toutes les valeurs susceptibles d'être extrêmes. En effet, la méthode BM extrait le maximum de chaque période définie au préalable (mois, année, etc.). Elle peut donc perdre certaines valeurs extrêmes qui pourraient survenir autour du maximum de la période (phénomène de cycles financiers ou *clustering*) alors qu'à la période suivante, le maximum pourrait être relativement faible. Au contraire, la méthode POT évite ce problème puisqu'elle extrait les maxima au-dessus d'un seuil fixé à l'avance. Cette méthode prend donc en compte les phénomènes de *clusters* ou grappes de volatilités caractéristiques

des rentabilités actifs financiers. Elle est donc particulièrement adaptée en Finance, lors que dans d'autres domaines, où les cycles sont absents, la méthode BM sera préférée.

Enfin, l'EMV a une robustesse et consistance certaines pour les séries qui nous concernent (pour lesquelles $\xi \geq -0.5$).

La justesse des résultats de cette modélisation s'observent sur la Figure 3.2.

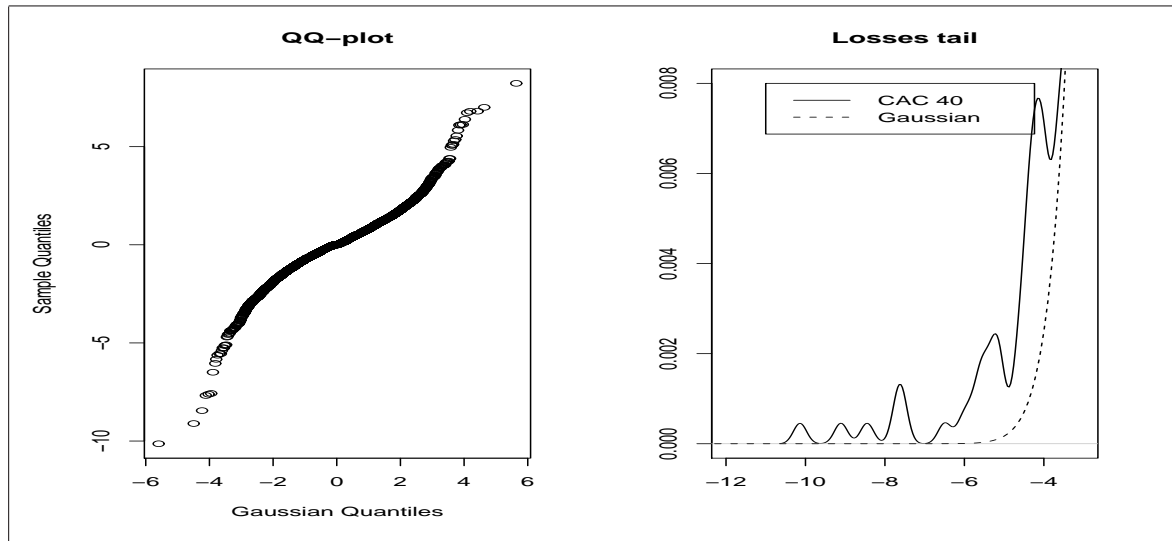


FIGURE 3.1 – QQ-plot et distribution de la queue des pertes du CAC 40 relativement à la loi Normale

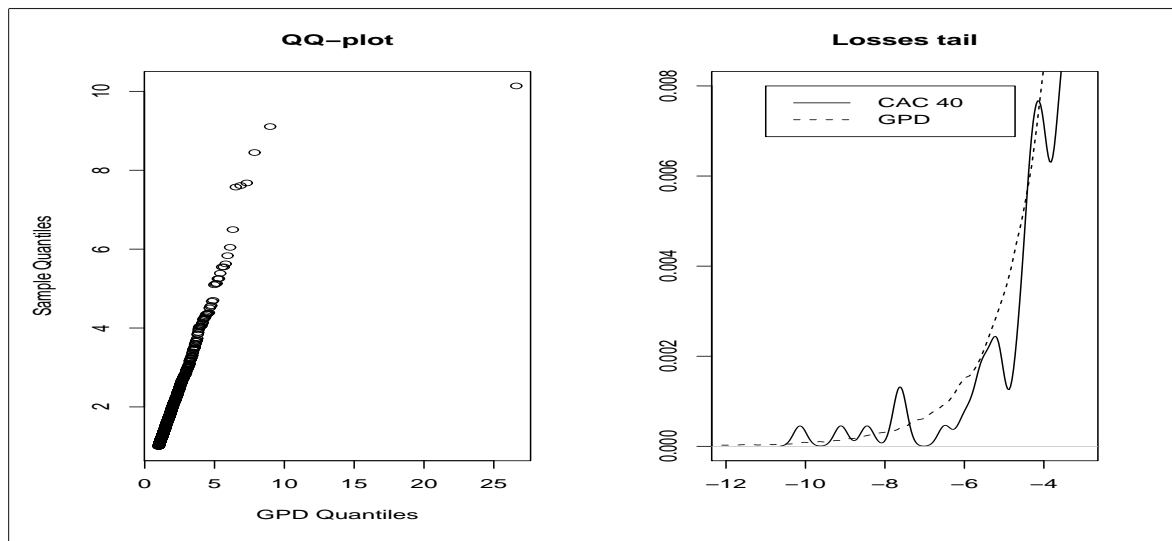


FIGURE 3.2 – QQ-plot et distribution de la queue des pertes du CAC 40 relativement à la loi GPD

Il faut ajouter une précision lorsque l'on calcule une VaR à partir de la TVE. Puisque la modélisation est ajustée à partir de l'échantillon des pertes et que nous voulons obtenir un quantile sur l'échantillon total, il nous faut réaliser une translation de la probabilité α pour calculer la VaR de notre échantillon.

Propriété 3.1 La probabilité d'occurrence α appliquée à la VaR calculée à l'aide d'une loi issue de la TVE

est égale à :

$$\alpha_{TVE} = 1 - \kappa * \alpha$$

avec κ le ratio entre le nombre de données de l'échantillon total et le nombre de données de l'échantillon des extrêmes (égal à la taille du bloc s pour la méthode BM).

Enfin, la modélisation TVE peut être combinée avec une modélisation dynamique des rentabilités financières dans le but d'améliorer les performances de la VaR paramétrique. Ceci est fait dans l'article de Frey et McNeil [9] via l'utilisation d'une modélisation ARMA-GARCH avec des résidus suivant une loi de la TVE.

Nous rappelons ici succinctement les tenants et aboutissants d'une telle modélisation.

3.3.2 Modèle dynamique

Nous avons, dans les parties précédentes, modélisé les rentabilités du portefeuille en faisant l'hypothèse que la moyenne et la volatilité étaient constantes dans le temps. Modéliser ces dernières de manière dynamique va nous permettre d'affiner la modélisation de la VaR. Nous nous rapprocherons encore un peu plus de la réalité des marchés financiers, notamment de la notion de *persistance* ou *clustering* des marchés (une forte volatilité est suivie d'une autre forte volatilité).

Nous supposons précédemment que les rentabilités d'un actif n'étaient pas auto-corrélées dans le temps. Levons maintenant cette hypothèse en supposant que le processus des rentabilités, que nous notons r_t , s'écrit :

$$r_t = \mu_t + \varepsilon_t$$

avec $\varepsilon_t \sim (0, \sigma_t)$ et iid. Ce qui change est le fait que les rentabilités ne vont pas avoir une moyenne et une volatilité constantes. Pour cela, nous allons spécifier un modèle autorégressif sur la moyenne conditionnelle, ainsi qu'un modèle hétéroscédastique, en modélisant le processus des volatilités conditionnelles.

Faisons un bref point théorique sur la modélisation hétéroscédastique². C'est Engel [7], qui, en 1982, pour modéliser le taux d'inflation au Royaume-Uni, suggéra un modèle ARCH (*AutoRegressive Conditional Heteroskedasticity*), où la variance conditionnelle en t est une fonction affine du carré du résidu du passé. On écrit les rentabilités sous la forme :

$$r_t = \mu_t + \sigma_t u_t$$

avec $u_t \sim (0, 1)$ et iid.

Et nous avons le modèle ARCH(p) suivant :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2$$

En 1986, Bollerslev [3] généralise l'approche ARCH(p) en introduisant le niveau atteint par la volatilité conditionnelle dans le passé. Il détermine le processus GARCH(p,q) de la façon suivante :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

Les conditions du modèle sont que $\alpha_i \geq 0, \beta_j \geq 0, \forall i$ et $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$ (condition de stationnarité de second ordre, c'est-à-dire variance non conditionnelle constante et finie).

2. Nous ne revenons pas sur la modélisation ARMA, ni sur son estimation.

Nous allons nous restreindre à l'étude du AR(1)-GARCH(1,1), le plus répandu en Finance, et qui se révèle souvent très performant. Il s'écrit :

$$\begin{aligned} r_t &= \mu_t + \sigma_t u_t \\ \mu_t &= \gamma + \theta r_{t-1} \\ \sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned} \quad (3.2)$$

Pour estimer les paramètres du GARCH(1,1), il faut déterminer la loi des innovations. Lorsque nous choisissons pour la distribution des innovations u_t du modèle (3.2) la loi normale centrée et réduite, nous avons, avec F_{t-1} l'information en $t - 1$:

$$r_t | F_{t-1} \sim \mathcal{N}(\mu_t, \sigma_t^2)$$

qui sont iid. Les paramètres du modèle GARCH peuvent alors être estimés en maximisant la Vraisemblance :

$$L(R; \omega, \alpha, \beta) = \prod_{t=1}^n \varphi(r_t | \mu_t, \sigma_t^2) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi}\sigma_t} \exp \left[-\frac{1}{2} \left(\frac{r_t - \mu_t}{\sigma_t} \right)^2 \right]$$

avec $\varphi(\cdot)$ la densité de la loi normale, $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 = \omega + \alpha (r_{t-1} - \mu_t)^2 + \beta \sigma_{t-1}^2$ et $R = (r_1, \dots, r_n)$.

En Finance, la véritable loi des innovations u_t est rarement conditionnellement normale. Ainsi nous utilisons une loi qui ne correspond pas à la loi utilisée pour calculer la Vraisemblance : nous réalisons une estimation par pseudo Maximum de Vraisemblance (PMV). White [16] et Gouriéroux, Monfort, Trognon [10] montrent que, sous diverses conditions de régularité, les estimateurs du PMV sont convergents et asymptotiquement normaux. Nous obtenons des estimations qui sont bonnes (estimateurs convergents), malgré le fait d'utiliser une hypothèse fautive (les innovations suivent une loi Normale).

Un test de Ljung-Box permet de tester l'autocorrélation de la série des résidus au carré du GARCH(1,1) pour savoir si l'hétéroscédasticité est entièrement modélisée. En outre, pour comparer différents modèles, les mesures AIC, BIC ou encore Likelihood Ratio sont très utiles.

Enfin, pour *filtrer* la volatilité à partir du GARCH(1,1), il faut initialiser la première valeur de la variance. Pour cela, nous conseillons d'utiliser la variance de long terme ω_0 déterminée à partir de :

$$\omega_0 = \frac{\omega}{1 - \alpha - \beta}$$

Nous pouvons maintenant modéliser les résidus du GARCH(1,1) (les innovations u_t) par une distribution issue de la TVE (cf. Frey et McNeil [9]).

3.3.3 Méthodologie du calcul d'une VaR TVE dynamique

La VaR prévue $t + 1$ se calcule à l'aide des prévisions déterminées à partir de l'information disponible en t . Elle est alors égale au quantile x_α de probabilité α des résidus u_t , quantile correspondant à la VaR des résidus, que l'on définit à partir des différentes modélisations étudiées auparavant, multiplié par la volatilité prévue en $t + 1$ par le GARCH auquel on ajoute la prévision de la moyenne en $t + 1$:

$$\text{VaR}(h, \alpha)_{t+1} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} x_\alpha = \hat{\gamma} + \hat{\theta} r_t + x_\alpha \sqrt{\hat{\omega} + \hat{\alpha} (r_t - \hat{\mu}_t)^2 + \hat{\beta} \hat{\sigma}_t^2}$$

CONCLUSION

Cet atelier a permis d'appliquer les méthodes statistiques issues de la Théorie des Valeurs Extrêmes à la modélisation et à la mesure du risque de marché en Finance. Nous avons alors calculé des Value-at-Risk robustes et ajustées à la réalité des marchés financiers.

Cette application est néanmoins transposable à d'autres champs d'étude qui sont nombreux à utiliser les résultats de la TVE : l'environnement, les assurances, la météorologie, etc.

BIBLIOGRAPHIE

- [1] BALKEMA, A. & de HAAN, L., 1974, *Residual Life Time at Great Age*, Annals of Probability, 2, 792-804.
- [2] BEIRLANT, J., GOEGBEUR, Y., SEGERS, J. & TEUGELS, J., 2004, *Statistics of Extremes - Theory and Applications*, Wiley, England.
- [3] BOLLERSLEV, T., 1986, *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, 31, 307-327.
- [4] BOLLERSLEV, T., 1987, *A conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return*, Review of Economics and Statistics, 69, 542-547.
- [5] CLAUSS, P., 2011, *Gestion de Portefeuille*, Dunod.
- [6] EMBRECHTS, P., KLUPPELBERG, C. & MIKOSCH, T., 1997, *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.
- [7] ENGEL, R., 1982, *Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation*, Econometrica, 50, 987-1007.
- [8] FISHER, R. & TIPPET, L., 1928, *Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample*, Proceedings of the Cambridge Philosophical Society, 24, 180-190.
- [9] FREY, R & MCNEIL, J., 2000, *Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series : an Extreme Value Approach*, Journal of Empirical Finance, 7, 271-300.
- [10] GOURIEROUX, C, MONFORT, A. & TROGNON, A., 1984, *Pseudo Maximum Likelihood Methods : Theory*, Econometrica, 52, 680-700.
- [11] HOSKING, J. & WALLIS, J., 1987, *Parameter and Quantile Estimation for the Generalized Pareto Distribution*, Technometrics, 29, 339-349.
- [12] PICKANDS, J., 1975, *Statistical Inference Using Extreme Order Statistics*, The Annals of Statistics, 3, 119-131.
- [13] ROOTZEN, H. & TAJVIDI, N., 1996, *Extreme Value Statistics and Wind Storm Losses : a Case Study*, Scandinavian Actuarial Journal, 70-94.
- [14] SHARPE, W., 1964, *Capital Asset Prices : a Theory of Market Equilibrium under Conditions of Risk*, Journal of Finance, 19, 425-442.
- [15] SMITH, R., 1989, *Extreme Value Analysis of Environmental Time Series : an Application to Trend Detection in Ground-Level Zone*, Statistical Science, 4, 367-393.
- [16] WHITE, H., 1982, *Maximum Likelihood Estimation of Misspecified Models*, Econometrica, 53, 1-16.