

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abderahamane Mira de Bejaia

Faculté des Sciences Exactes
Département de Mathématiques



Mémoire de fin de cycle

En vue de l'obtention d'un Master académique en Mathématiques
Option : Probabilités Statistique et Applications

Thème

Estimation de quantiles et application en finance.

Présenté Par :

M^{lle} BENBOUDJEMA Mouna

Devant le Jury composé de :

<i>M^{me}</i>	K.TIMERIDJINE	Présidente	M.C.A	U.A/Mira, Bejaia
<i>M^{me}</i>	N. SAADI	Promotrice	M.C.A	U.A/Mira, Bejaia
<i>M^{me}</i>	A. BARECHE	Examinatrice	Professeur	U.A/Mira, Bejaia
<i>M^{me}</i>	S.CHEKKAL	Examinatrice	Doctorante	U.A/Mira, Bejaia

Année universitaire : 2019 / 2020

Remerciements

Au terme de mon modeste travail, je tiens à remercier tous d'abord Dieu, le tout puissant de m'avoir accordé la volonté, force, santé, détermination et le courage afin d'accomplir mon mémoire.

*Je tiens à exprimer ma profonde gratitude à ma promotrice M^{me}. **SAADI** pour son encadrement. Je suis également reconnaissante pour la confiance qu'elle m'a accordée, sa disponibilité et son engagement. Ses remarques constructives m'ont beaucoup apporté tout au long de ce travail.*

*Je tiens vivement à remercier M^{me}. **TIMERIDJINE** pour l'immense honneur qu'elle m'a fait en décidant de présider le jury de ce mémoire.*

*Mes remerciements chaleureux s'adresse également à M^{me}. **BARECHE** et M^{me}. **CHEKKAL** d'avoir accepté d'examiner et d'évaluer ce modeste travail.*

Je remercie tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste mémoire.

Enfin, je tiens à remercier ma famille et particulièrement mes parents pour leur affection et leur soutien tout au long de mes études.

Dédicaces

Ce modeste travail est dédié :

À mes chers parents pour leurs soutient et leurs prières.

À mes frères et mes sœurs.

À ma famille.

À mes amis.

À tous ceux qui me sont chers...

Table des matières

Table des figures	i
Liste des tableaux	ii
Notations et Abréviations	iii
Introduction générale	3
1 Quelques éléments de la théorie des valeurs extrêmes	5
1.1 Introduction	5
1.2 Statistique d'ordre et définitions	5
1.2.1 Fonction de répartition empirique d'un échantillon	6
1.2.2 lois des valeurs extrêmes	7
1.2.3 Densité conjointe de n statistiques d'ordre	8
1.3 Quantile	9
1.3.1 Fonction quantile de queue	10
1.3.2 Quantile extrême	10
1.4 Loi généralisée des valeurs extrême et domaines d'attraction	10
1.4.1 Loi des valeurs extrêmes	10
1.4.2 Distribution des excès (P.O.T)	12
1.5 Caractérisation des domaines d'attraction	15
1.5.1 Quelques résultats sur les fonctions à variations régulières	15
1.6 Conclusion	17
2 Estimation de quantiles	18
2.1 Introduction	18
2.2 Approche paramétrique	18
2.3 Approche des quantiles extrêmes par la loi des valeurs extrêmes	19
2.4 Approche des quantiles extrêmes par la méthode des excès P.O.T	21
2.4.1 Principe de la méthode	21
2.4.2 La detection du seuil	21
2.5 Approche semi-paramétrique	23

2.5.1	Lois à queues lourdes	23
2.5.2	Estimation du paramètre de la loi des valeurs extrêmes	23
2.6	La méthode non paramétrique	24
2.6.1	Estimation non paramétrique de la densité de probabilité par la méthode du noyau	25
2.6.2	Estimation de la fonction de répartition	26
2.6.3	Choix du paramètre de lissage	27
2.6.4	Estimation non paramétrique d'un quantile	29
2.7	Conclusion	30
3	Valeur à Risque et estimation	31
3.1	Introduction	31
3.2	Les mesures de risque	31
3.2.1	Caractéristiques d'une mesure de risque	32
3.2.2	Mesure de risque cohérente	32
3.2.3	Mesure de risque monétaire	32
3.3	Définitions	32
3.4	Valeur à Risque (Value at risk) VaR	33
3.4.1	Interprétation mathématique	34
3.4.2	Propriétés de la VaR	35
3.5	Estimation de la VaR	35
3.5.1	La méthode paramétrique	35
3.5.2	La méthode historique	36
3.5.3	La méthode des excès	36
3.6	Conclusion	36
4	Application numérique	37
4.1	Introduction	37
4.2	Type des données	38
4.3	Analyse descriptive des prix de clôture	39
4.4	Calcul des rendements journaliers	41
4.5	La VaR par la méthode paramétrique	45
4.6	La VaR historique	46
4.7	La VaR par la théorie des valeurs extrêmes	47
4.7.1	Détection du seuil	47
4.8	Comparaison	51
	Conclusion générale	52
	Bibliographie	56
	Annexe	57

Table des figures

1.1	Dépassement de seuil (POT).	13
1.2	Densité des lois de valeurs extrêmes, avec $\xi = 0.5$ pour la loi de Pareto, $\xi = 0$ pour la loi exponentielle et $\xi = -0.5$ pour la loi de Pareto II.	14
4.1	Graphique des prix de clôture entre Avril 2018 - Avril 2020.	39
4.2	l'histogramme des prix de clôture	40
4.3	Rendements logarithmiques journaliers du KVV.	42
4.4	Histogramme des variations des prix	43
4.5	Histogramme des variations des prix	45
4.6	La distribution moyenne des excès.	48
4.7	L'approximation des excès par une GPD.	49
4.8	QQ-plot.	50
4.9	Les prix de clôture de l'action KvW en Euros.	58

Liste des tableaux

2.1	Noyaux usuels.	26
4.1	Déscription de la serie des prix de clôture.	39
4.2	Déscription de la serie des variations des prix.	41
4.3	Les paramètres obtenus	49
4.4	La VaR pour les différentes méthodes de niveau 0.95	51

Notations et Abréviations

Abréviations

TVE	Théorie des Valeur Extrêmes.
i.i.d	Independant et Identiquement Distirbué.
GEV	Generalized Extreme Value(Distribution des Valeurs Extrêmes Généralisée).
GPD	Generalized Pareto Distribution(Distribution de Pareto Généralisée).
POT	L'approche par dépassement de seuil.
EMV	Méthode de Maximum de Vraisemblance.
MSE	Mean Square Error(Erreur Quadratique Moyenne).
MISE	Mean integrated square error(L'erreur Quadratique Moyenne Intégre).
AMISE	L'erreur Quadratique Moyenne Intégre Asymptotique.
ISE	Integrated Squared Error(L'erreur Quadratique Intégre).
UCV	Unbaised Cross Validation.
BCV	Baised Cross Validation.
VaR	Value at Risk (Valeur à Risque).

Notations

F	Fonction de répartition.
F_n	Fonction de répartition empirique.
F^{\leftarrow}	Inverse généralisée de la fonction de répartition.
F^{-1}	Inverse de la fonction de répartition.
x_F	Point terminal.
\mathbb{R}	Ensemble des réels.
D	Demaine d'attraction.
Θ	Espace des paramtres.
$\xrightarrow{p-s}$	Convergence presque sûre.

Introduction générale

La théorie des valeurs extrêmes est une branche de la statistique qui convient le mieux au calcul de risque dans le cadre de la gestion des marchés financiers, elle fournit des modèles statistiques pour étudier les événements rares. Elle se repose principalement sur des distributions limites des extrêmes et leurs domaines d'attraction. Cependant, on y retrouve deux modèles : loi généralisée des extrêmes (GEV : "Generalized Extreme Value") et loi de Paréto généralisée (GPD : "Generalized Pareto Distribution").

L'intérêt et l'attention portée sur la théorie des valeurs extrêmes ne fait qu'augmenter durant ces dernières années, et cela sur les deux plans théorique et pratique. Les domaines d'applications sont en effet très variés, on se doit d'en citer quelques uns : hydrologie, météorologie, biologie, ingénierie, gestion de l'environnement, finance, assurance, sciences sociales, etc [34]. Aujourd'hui, la gestion des risques est devenue une pratique fondamentale dans tous ces domaines. Dans ce travail, on s'intéresse aux domaines financiers. À titre d'exemple, nous mentionnons la crise de 1929, le crash de 1987 et la crise de 2007-2008. Ce sont des événements rares [38], qui ont une faible probabilité d'apparition comparée à la probabilité des événements qui se produisent le plus souvent. Ces événements ont un coût économique énorme. Ils sont dits extrêmes quand il s'agit des valeurs beaucoup plus grandes ou plus petites que celle desservies habituellement.

L'une des problématiques majeures en gestion financière est la détermination du risque financier d'un portefeuille d'actifs. En 1993, le Groupe des Trente [2], composé de banquiers, d'autorités de contrôle et d'académiciens, a publié un document qui recommandait notamment l'usage d'une nouvelle mesure de risque nommée Valeur à Risque (Value at Risk en anglais) comme critère de mesure pour le risque de marché. En termes statistiques, la Valeur à Risque est un quantile extrême de la distribution de la variable aléatoire d'intérêt.

L'objectif de ce travail est l'estimation de quantile extrême ou la Valeur à Risque et d'appliquer les différentes méthodes de calcul de ces deux derniers en finance, et de comparer les différents estimateurs.

Ce travail s'articule autour de quatre chapitres et une introduction générale, le plan de travail est le suivant :

Dans le chapitre 1, nous rappelons quelques éléments théoriques essentiels de la théorie des valeurs extrêmes (TVE). Il contient des rappels sur la statistique d'ordre et les définitions du quantile. On introduit le théorème de Fisher et Tippett qui sert à modéliser les maximums prouvés en 1928 [14] et Gnedenko en 1943 [19]. On conclut ce chapitre par la notion de fonction à variation régulière.

Le chapitre 2 présente les différentes méthodes statistiques et techniques d'estimation du quantile et de l'indice des valeurs extrêmes.

Le troisième chapitre, consiste à introduire une mesure de risque VaR (Value at Risk) c'est le quantile extrême. Pour cela, on commence par la définition de la VaR ainsi que les différentes étapes associées à son calcul. Puis, on cite plusieurs méthodes d'estimations qui sont basées sur la méthode historique, paramétrique et la méthode semi-paramétrique.

Dans le quatrième chapitre, on présente une application sur une action boursière parmi d'autres actions qu'on trouve en finance. Dans le but d'estimer la Valeur à Risque et le quantile extrême en faisant appel à des fonctions et des méthodes prédéfinies sous le logiciel *R*.

Enfin, on conclut ce travail par une conclusion générale.

Chapitre 1

Quelques éléments de la théorie des valeurs extrêmes

1.1 Introduction

Les quantiles sont fréquemment utilisés en Statistique. Par exemple, la médiane est un indicateur robuste de la tendance centrale d'une population, l'intervalle interquartile est un bon indicateur de sa dispersion. Dans la pratique, les domaines d'utilisation des quantiles sont assez variés. En biologie, Gannoun et al [17] utilisent les quantiles conditionnels pour estimer des courbes de référence permettant d'analyser certaines propriétés biophysiques de la peau. Les quantiles représentent également un moyen robuste de prévision (voir par exemple De Gooijer et al [8] et Gannoun et al [18]). En pratique, ces quantiles sont calculés suivant un critère d'ordre sur les observations. La théorie des valeurs extrêmes a pour but d'étudier et de caractériser le comportement des valeurs extrêmes d'un échantillon de variables aléatoires. On souhaite estimer des petites probabilités ou des quantités dont la probabilité d'observation est très faible, c'est-à-dire proche de zéro. Ces quantités sont appelées quantiles, on parle de quantile extrême lorsque l'ordre du quantile (probabilité d'observation) converge vers zéro quand la taille de l'échantillon tend vers l'infini. Ces quantiles extrêmes se situent dans les queues de distributions des lois de probabilité. Dans ce chapitre, nous donnons des définitions et des résultats sur la théorie des valeurs extrêmes dans le cas univarié réel.

1.2 Statistique d'ordre et définitions

Définition 1.2.1. (Fonction de queue). Soit X une variables aléatoire de densité f_X et de fonction de répartition F_X . On appelle fonction de survie (de queue) la fonction définie par

$$\bar{F}_X = 1 - F_X(x).$$

Définition 1.2.2. (Point terminal). Le point terminal d'une distribution F_X est défini par :

$$x_F = \sup\{x \in \mathbb{R} : F_X(x) < 1\}.$$

Définition 1.2.3. (Inverse généralisé). On appelle inverse généralisé de F , l'application notée F^{\leftarrow} définie par :

$$F^{\leftarrow}(p) = \inf\{x : F(x) \geq p\}, p \in [0, 1].$$

L'inverse généralisé F^{\leftarrow} coïncide avec l'inverse F^{-1} lorsque la fonction F est strictement croissante et continue.

Définition 1.1. (Distribution d'un échantillon). Soit (X_1, X_2, \dots, X_n) un échantillon aléatoire issu d'une variable aléatoire X de densité de probabilité f_X et de fonction de répartition F_X . La vraisemblance de l'échantillon (la distribution du n-échantillon) (X_1, X_2, \dots, X_n) au point $x = (x_1, \dots, x_n)$ est :

$$L_X(x_1, x_2, \dots, x_n) = L_X(x) = f_{X_1}(x_1)f_{X_2}(x_2)\dots f_{X_n}(x_n) = \prod_{i=1}^n f_{X_i}(x_i) = (f_X(x_i))^n.$$

1.2.1 Fonction de répartition empirique d'un échantillon

Définition 1.2. Soit X_1, X_2, \dots, X_n un échantillon aléatoire issu d'une variable aléatoire X . Désignons par $F_n(x)$ la proportion des n variables X_1, X_2, \dots, X_n qui sont inférieures à x . $F_n(x)$ est donc une variable aléatoire pour tout x qui définit ainsi une fonction aléatoire appelée fonction de répartition empirique de l'échantillon, dont la réalisation sont des fonctions en escalier de sauts égaux à $\frac{1}{n}$. Si les x_i sont ordonnées par valeurs croissantes alors :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{(X_i \leq x)} = \begin{cases} 0, & \text{si } x < x_1; \\ \frac{1}{n}, & \text{si } x_1 \leq x < x_2; \\ \vdots & \\ \frac{i-1}{n}, & \text{si } x_{i-1} \leq x < x_i; \\ \vdots & \\ 1, & \text{si } x \geq x_n. \end{cases}$$

Propriétés asymptotiques de la fonction de répartition empirique

Théorème 1.2.1. Pour tout x , on a

$$F_n(x) \xrightarrow{P.s} F(x).$$

Théorème 1.2.2. [10](*Glivenko-Cantelli*)

La convergence de F_n vers F est presque sûrement uniforme, c'est-à-dire que :

$$D_n = \sup_x |F_n(x) - F(x)| \rightarrow 0.$$

Théorème 1.2.3. [10](*Kolmogorov*)

$$\lim_{n \rightarrow \infty} P(\sqrt{n}D_n < y) = \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2 y^2).$$

Ce théorème signifie que la distribution asymptotique de la variable aléatoire D_n est connue et ne dépend pas de la variable de départ X , et permet de calculer des limites pour les valeur de D_n . La loi exacte de la variable D_n a été tabulée.

1.2.2 lois des valeurs extrêmes

Définition 1.3. (Statistique d'ordre) Soit X_1, \dots, X_n un n-échantillon aléatoire issu d'une variable aléatoire X . Notons par $X_{(1,n)} \leq X_{(2,n)} \leq \dots \leq X_{(n,n)}$ le réarrangement croissant de l'échantillon $\{X_i, i = 1, \dots, n\}$. Pour tout $i = 1, \dots, n$, la variable aléatoire $X_{(i,n)}, i = \overline{1, n}$ s'appelle la i -ème statistique d'ordre de l'échantillon.

Lois de $X_{(1,n)}$ et de $X_{(n,n)}$

Soit F (resp. f) la fonction de répartition (resp. la densité) de la variable aléatoire X . Soit $F_{k,n}$ (resp. $f_{k,n}$) les fonctions de répartition (resp. les densités) de la k -ième statistique d'ordre $X_{(k,n)}$. Deux statistiques d'ordre sont particulièrement intéressantes pour l'étude des événements extrêmes. Ce sont les statistiques d'ordre extrême qui correspondent à la plus petite statistique d'ordre $X_{(1,n)}$ (ou statistique du minimum)

$$X_{(1,n)} = \min_{1 \leq i \leq n} X_i;$$

et à la plus grande statistique d'ordre $X_{(n,n)}$ (ou statistique du maximum)

$$X_{(n,n)} = \max_{1 \leq i \leq n} X_i.$$

A. Loi de $X_{(1,n)} = \min_{1 \leq i \leq n} X_i$.

On a

$$\begin{aligned}
 F_{(1,n)}(x) &= \mathbb{P}(X_{(1,n)} < x) \\
 &= 1 - \mathbb{P}(X_{(1,n)} \geq x) \\
 &= 1 - \mathbb{P}\left(\min_{1 \leq i \leq n} X_i \geq x\right) \\
 &= 1 - \prod_{i=1}^n \mathbb{P}(X_i \geq x) = 1 - \prod_{i=1}^n [1 - \mathbb{P}(X_i < x)] \\
 &= 1 - \prod_{i=1}^n [1 - F(x)]
 \end{aligned}$$

donc

$$\begin{aligned}
 F_{(1,n)}(x) &= 1 - [1 - F(x)]^n. \\
 f_{(1,n)}(x) &= n [1 - F(x)]^{n-1} f(x).
 \end{aligned}$$

B. Loi de $Y_{(n,n)} = \max_{1 \leq i \leq n} X_i$

$$\begin{aligned}
 F_{(n,n)}(x) &= \mathbb{P}(X_{(n,n)} < x) \\
 &= \mathbb{P}(\max_{1 \leq i \leq n} X_i < x) \\
 &= \prod_{i=1}^n \mathbb{P}(X_i < x) \\
 &= \prod_{i=1}^n [F(x)] \\
 &= [F(x)]^n
 \end{aligned}$$

et

$$f_{(n,n)}(x) = nF(x)^{n-1}f(x).$$

1.2.3 Densité conjointe de n statistiques d'ordre

Lemme 1.2.4. Soit X_1, \dots, X_n n variables aléatoires (i.i.d) issues d'une variable aléatoire X de fonction de répartition F_X et de densité f_X continue, alors la densité de probabilité de $(X_{(1,n)}, \dots, X_{(n,n)})$ est donnée par

$$f_{(X_{(1,n)}, \dots, X_{(n,n)})}(x_1, \dots, x_n) = n! \prod_{i=1}^n f_X(x_i), \quad x_1 \leq x_2 \leq \dots \leq x_n.$$

Distribution conjointe de la k-ième statistiques d'ordre

Lemme 1.2.5. L'expression de la distribution de $X_{(k,n)}$ est :

$$\begin{aligned} F_{(k,n)}(x) &= \mathbb{P}(X_{(k,n)} \leq x) \\ &= \sum_{r=k}^n C_n^r [F_X(x)]^r [1 - F_X(x)]^{n-r}. \end{aligned}$$

Nous en déduisons que la fonction de densité est :

$$f_{(k,n)}(x) = \frac{n!}{(k-1)!(n-k)!} [F_X(x)]^{k-1} [1 - F_X(x)]^{n-k} f_X(x).$$

1.3 Quantile

Définition 1.3.1. (Quantile). Pour une variable $X \in \mathbb{R}$, la fonction quantile se définit à partir de l'inverse de sa fonction de répartition. Quand cette fonction de répartition est strictement croissante, son inverse est défini sans ambiguïté. Mais une fonction de répartition reste constante sur tout intervalle dans lequel la variable aléatoire ne peut pas prendre de valeurs. De manière générale, si $F(\cdot)$ désigne la fonction de répartition de la variable X , on appelle fonction quantile de X la fonction qui, à $\alpha \in [0, 1]$, associe

$$q(\alpha) = x_\alpha = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}, \alpha \in [0, 1]. \quad (1.1)$$

Si F est strictement croissante et continue, alors x_α est l'unique nombre réel tel que :

$$F(x_\alpha) = \alpha.$$

Le quantile en tant que racine d'une équation

Soit $\alpha \in [0, 1]$, posons $u = 2\alpha - 1$. On introduit une nouvelle fonction quantile notée $q(\cdot)$ définie sur l'intervalle $[-1, 1]$ par :

$$q(u) = F^{-1}\left(\frac{1+u}{2}\right), -1 \leq u \leq 1. \quad (1.2)$$

Nous remarquons que, le quantile est indexé par $u \in [-1, 1]$. La définition de la fonction quantile $q(\cdot)$ donnée par (1.2) nous donne, à l'aide du signe (resp. la valeur absolue) de u , une idée sur l'orientation (resp. l'ordre) du quantile par rapport à la médiane. En effet :

- Pour $u = 0$, $q(0)$ est la médiane (le quantile d'ordre $\alpha = \frac{1}{2}$),
- Si u est négatif (resp. positif), le quantile d'ordre u est à gauche (resp. à droite) de la médiane.
- Si $|u|$ est proche de 0, le quantile correspondant est proche de la médiane (quantile d'ordre $\alpha = \frac{1}{2}$); si $|u|$ est proche de 1, le quantile correspondant est un quantile "extrême" (quantile d'ordre α proche de 0 ou de 1).

1.3.1 Fonction quantile de queue

La fonction quantile de queue est définie par :

$$U(t) = F^{\leftarrow} \left(1 - \frac{1}{t} \right) \quad \text{avec} \quad 1 < t < \infty.$$

1.3.2 Quantile extrême

Définition 1.3.2. On appelle quantile extrême le quantile d'ordre $(1 - \alpha)$, défini par :

$$\begin{aligned} q(1 - \alpha) = x_{1-\alpha} &= \inf \{ x \in \mathbb{R} : F(x) \geq 1 - \alpha \} \\ &= F^{-1}(1 - \alpha). \end{aligned}$$

Où α proche de zéro.

Quantile empirique

La fonction de quantile empirique de l'échantillon X_1, \dots, X_n , est donnée par :

$$q_n(\alpha) = F_n^{-1}(\alpha) = \inf \{ x \in \mathbb{R} : F_n(x) \geq \alpha \}, \quad 0 < \alpha < 1.$$

La fonction quantile empirique de la queue correspondante est définie par :

$$U_n(t) = F_n^{\leftarrow} \left(1 - \frac{1}{t} \right) \quad \text{avec} \quad 1 < t < \infty.$$

1.4 Loi généralisée des valeurs extrême et domaines d'attraction

1.4.1 Loi des valeurs extrêmes

Le principal résultat de la théorie des valeurs extrêmes repose sur le Théorème de Fisher et Tippett (1928)[14]. Le résultat établit c'est la loi asymptotique du maximum $X_{n,n} = \max(X_1, \dots, X_n)$ d'un échantillon de n variables. Il a été démontré notamment par Gnedenko [25].

Théorème 1.4.1. (Fisher-Tippett)[14]. *Sous certaines conditions de régularité sur la fonction de répartition F , il existe un paramètre réel ξ et deux suites $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ et une suite non dégénérée H_ξ telle que :*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{X_{n,n} - b_n}{a_n} \leq x \right] = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H_\xi(x) \quad (1.3)$$

avec

$$H_\xi(x) = \begin{cases} \exp \left[(-x)^{-\frac{1}{\xi}} \right] & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \quad \text{et } \xi > 0, \quad \text{Loi de Frechet.}$$

$$H_\xi(x) = \begin{cases} \exp \left[-(-x)^{-\frac{1}{\xi}} \right] & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases} \quad \text{et } \xi < 0, \quad \text{Loi de Weibull.}$$

$$H_\xi(x) = \exp[-\exp(-x)] \text{ si } x \in \mathbb{R} \text{ et } \xi = 0 \quad \text{Loi de Gumbel.}$$

et H_ξ la fonction de répartition de la loi des valeurs extrêmes (EVD).

Remarque 1.4.1. [34].

Ce théorème donne un résultat très intéressant : quelle que soit la loi limite de la variable parente, la loi limite des extrêmes a toujours la même forme. Bien que le comportement de ces lois soit complètement différent, elles peuvent être combinées en une seule paramétrisation contenant un unique paramètre ξ qui contrôle la "lourdeur" de la queue de loi appelé indice des valeurs extrêmes (ou indice de queue).

Définition 1.4.1. :Distribution généralisée des valeurs extrêmes (GEV). Une distribution est dite dans la loi des valeurs extrêmes généralisée (Generalized Extreme Value), que l'on notera *GEV*, si et seulement si elle est non-dégénérée et si sa fonction de répartition est de la forme

$$H_\xi(x) = \begin{cases} \exp \left(-(1 + \xi x)^{-\frac{1}{\xi}} \right), & \xi \neq 0, \quad 1 + \xi x > 0 \\ \exp[-\exp(-x)], & \xi = 0, \quad -\infty \leq x \leq +\infty \end{cases} \quad (1.4)$$

En introduisant les paramètres de localisation μ et de dispersion σ dans la paramétrisation (1.4), on obtient la forme la plus générale de la GEV :

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

Définition 1.4.2. La densité GEV de la loi des valeurs extrêmes s'en déduit par simple dérivation. Donc elle est donnée par :

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

Remarque 1.4.2. La valeur prise par le paramètre de forme ξ définit trois lois connues dans la littérature :

- le cas $\xi = 0$ correspond à la loi Gumbel "queue moyenne" (bounded). On dit que $H(\cdot)$ appartient au domaine d'attraction de Gumbel qui regroupe les lois ayant une fonction de survie à décroissance exponentielle. C'est le cas des lois normale, gamma, exponentielle, etc.
- le cas $\xi > 0$ correspond à la loi de Fréchet "queue lourde" (heavy tail) [16]. on dit que $H(\cdot)$ appartient au domaine d'attraction de Fréchet. Il contient les lois dont la fonction de survie décroît comme une fonction puissance. On parle aussi de lois à queue lourde. Dans ce domaine d'attraction, on trouve les lois de Pareto, de Student, de Cauchy, etc.
- le cas $\xi < 0$ correspond à la loi de Weibull "queue légère" (light tail). la fonction de répartition $H(\cdot)$ appartient au domaine d'attraction de Weibull. Ce domaine contient les lois dont le point terminal x_H est fini. C'est le cas par exemple des lois uniformes, lois beta, etc.

1.4.2 Distribution des excès (P.O.T)

Principe de la méthode

L'approche par dépassements de seuil, en anglais "Peaks-Over-Threshold approach" notée **POT**, repose sur l'utilisation des statistiques d'ordre supérieur de l'échantillon. Elle consiste à ne conserver que les observations dépassant un certain seuil. L'excès au-delà du seuil est défini comme l'écart entre l'observation et le seuil. Cette méthode initialement développée par Pickands (1975) [32] et abondamment étudiée par divers auteurs tels que de Smith (1987) [42], Davison et Smith (1990) [7], ou Reiss et Thomas (2001) pour de plus amples références, repose sur l'utilisation des statistique d'ordre supérieurs de l'échantillon [12].

Plus précisément, soit un échantillon de variables aléatoires indépendantes, identiquement distribuées (*i.i.d*) X_1, X_2, \dots, X_n de fonction de répartition F et de point terminal x_F . Soit u un seuil fixé (non aléatoire) tel que $u < x_F$. Considérons les N_u observation X_{i1}, \dots, X_{N_u} dépassant le seuil u tel que N_u suit la loi se poisson. On appelle excès au-delà du seuil u les $Y_j = X_{ij} - u$, où $j = 1, \dots, N_u$, l'excès au dela de seuil u . voire figure 1.1 ci-dessous [1] :

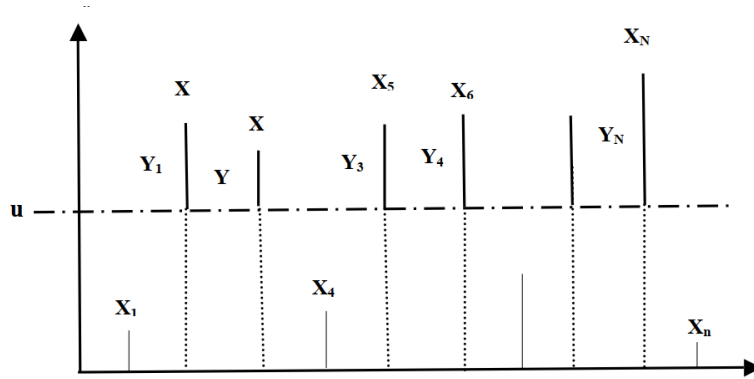


FIGURE 1.1 – Dépassement de seuil (POT).

On notera F_u la fonction de répartition de l'excès Y au delà du seuil u . La loi des excès est celle de variables aléatoires *i.i.d.* admettant pour fonction de répartition :

$$F_u(x) = \mathbb{P}(Y \leq x / X > u).$$

Représentant la probabilité que la variable aléatoire X ne dépasse pas le seuil u de au moins une quantité x sachant qu'elle dépasse u . F_u décrit ainsi la loi de Y sachant que $X > u$. On peut la réécrire en fonction de F à l'aide du résultat suivant.

On a pour $x \geq 0$:

$$F_u(x) = \mathbb{P}(X - u \leq x / X > u) = \frac{F(x+u) - F(u)}{1 - F(u)} = 1 - \frac{\bar{F}(x+u)}{\bar{F}(u)} \quad (1.5)$$

où $\bar{F} = 1 - F$ est la fonction de survie. On a alors le résultat suivant :

Théorème 1.4.2. *Balkema et Haan, Pickands [4].*

Si F appartient à l'un des trois domaines d'attraction de la loi des valeurs extrêmes, (Fréchet, Gumbel ou Weibull), alors il existe une fonction $\sigma(u) > 0$ strictement positive et un $\xi \in \mathbb{R}$ tel que la loi des excès F_u peut être uniformément approchée par une loi de Pareto généralisée notée $G_{\xi, \sigma(u)}$, i.e.

$$\lim_{u \rightarrow x_F} \sup_{x \in]0, x_F - u[} |F_u(x) - G_{\xi, \sigma(u)}(x)| = 0 \quad (1.6)$$

Avec : x_F est le point terminal et $G_{\xi, \sigma(u)}(x)$ est la fonction de répartition de la loi de Pareto Généralisée.

tel que :

$$G_{\xi, \sigma}(x) = \begin{cases} 1 - (1 + \xi \frac{x}{\sigma})^{-1/\xi} & \text{si } \xi \geq 0, \\ 1 - \exp(-\frac{x}{\sigma}), & \text{si } \xi = 0 \text{ et } x > 0. \end{cases}$$

Où : $x > 0$ pour $\xi \geq 0$. et $0 \leq x \leq -\frac{\sigma}{\xi}$ pour $\xi < 0$.

Remarque 1.4.3. Ce théorème montre l'existence d'une relation étroite entre la GPD et la GEV. Pickands (1975)[32] a montré que pour n'importe quelle loi F , l'approximation GPD définie par (1.6) n'est vérifiée que s'il existe des constantes de normalisation et une loi non dégénérée telle que le résultat donné par (1.3) est vérifié. Dans ce cas, si H est écrite sous la forme d'une GEV (équation 1.4) alors l'indice de queue ξ est le même que celui de la GPD.

De même pour la GPD, le cas où $\xi > 0$ correspond aux lois à queues épaisses, pour lequel $1 - G$ se comporte comme une puissance, $x^{-\frac{1}{\xi}}$ pour x assez élevé. Si $\xi = 0$, on trouve $1 - \exp(-\frac{x}{\sigma})$, c'est la loi exponentielle de paramètre σ et enfin pour $\xi < 0$ c'est la loi de Pareto de type II à support borné (figure 1.2)[34].

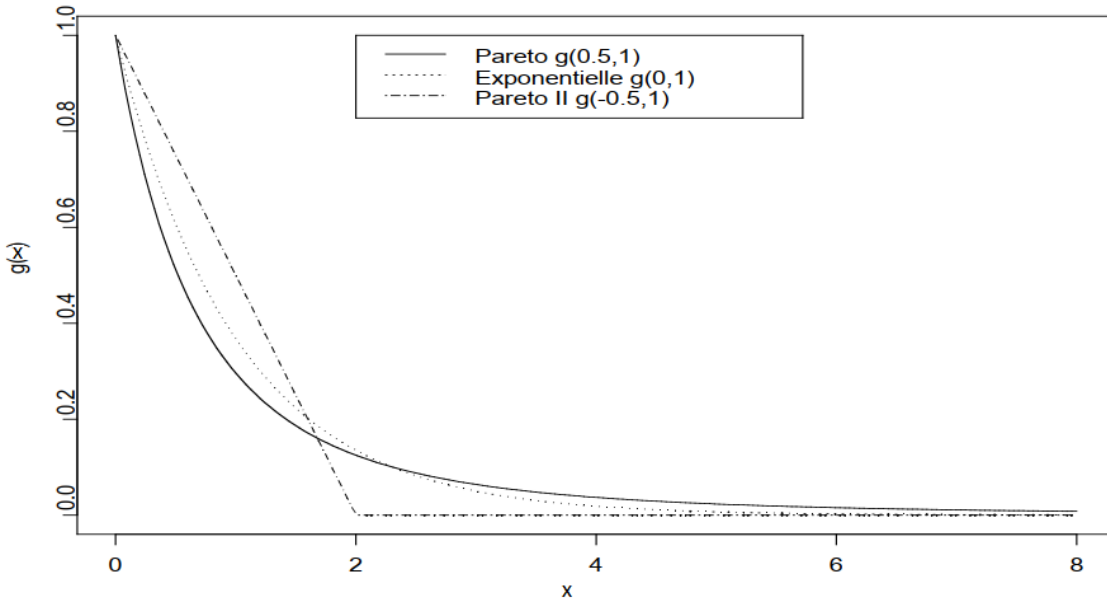


FIGURE 1.2 – Densité des lois de valeurs extrêmes, avec $\xi = 0.5$ pour la loi de Pareto, $\xi = 0$ pour la loi exponentielle et $\xi = -0.5$ pour la loi de Pareto II.

La densité GPD de la loi des valeurs extrêmes s'en déduit par simple dérivation. Donc elle est donnée par :

$$g_{GPD}(x) = \begin{cases} \frac{1}{\sigma} (1 + \xi \frac{x}{\sigma})^{-\frac{1}{\xi}-1} & \text{si } \xi \neq 0, \\ \frac{1}{\sigma} \exp(-\frac{x}{\sigma}), & \text{si } \xi = 0. \end{cases}$$

Le paramètre ξ est le même d'une GEV .

ξ : paramètre de forme (queue).

σ : paramètre d'échelle.

Remarque 1.4.4. La loi GPD présente quelques particularités.

- Si $\sigma = 1$, on parle la loi GPD standard.
- Si $\xi = 0$, la GPD correspond à une loi exponentielle d'espérance σ .
- Si $\xi = -1$, elle correspond à une loi uniforme sur $[0, \sigma]$.
- Si $\xi > 0$, on retrouve la loi de Pareto décentrée.
- Si $\xi < 0$, nous obtenons la loi de Pareto de type II.

Propriétés de la Distribuon de Paréto Généralisée (GPD)[34].

$$E(X) = \frac{\sigma}{1-\xi} \text{ avec } (\xi < 1)$$

$$V(X) = \frac{\sigma^2}{(1-\xi)^2(1-2\xi)^2} \text{ avec } (\xi < \frac{1}{2})$$

1.5 Caractérisation des domaines d'attraction

Nous allons donner par la suite, des conditions sur la fonction de répartition F pour qu'elle appartienne à l'une des trois domaines d'attraction (Gumbel, Weibull et Fréchet).

1.5.1 Quelques résultats sur les fonctions à variations régulières

Pour décrire plus en détail les fonctions à variations régulières, il est nécessaire de commencer par une définition des fonctions à variations lentes.

Définition 1.5.1. [25]. On dit qu'une fonction l est à variations lentes à l'infini si $l(x) > 0$ pour x assez grand et si pour tout $\lambda > 0$, on a

$$\lim_{x \rightarrow \infty} \frac{l(\lambda x)}{l(x)} = 1$$

Définition 1.5.2. On dit qu'une fonction f est à variations régulières d'indice $\alpha \in \mathbb{R}$ à l'infini si f est positive à l'infini (i.e s'il existe A tel que pour tout $x \geq A$, $f(x) > 0$) et on note $f \in RV_\alpha$ si pour tout $\lambda > 0$

$$\lim_{t \rightarrow \infty} \frac{f(\lambda t)}{f(t)} = \lambda^\alpha$$

avec RV : Variation régulière.

A. Domaine d'attraction de Fréchet

Théorème 1.5.1. [19]. Une fonction de répartition F appartient au domaine d'attraction Φ_α (de Fréchet avec un indice des valeurs extrêmes $\xi > 0$) si et seulement si $x_F = \infty$ et si la fonction de survie $\bar{F} \in RV_{-1/\xi}$ c'est à dire :

$$\forall x > 0, \lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-1/\xi}.$$

Autrement dit, une fonction de répartition F appartenant au domaine d'attraction de Fréchet s'écrit sous la forme :

$$F(x) = 1 - x^{-1/\xi}l(x) \quad \text{avec } l \in RV_0.$$

Les suites de normalisation (a_n) et (b_n) sont données dans ce cas pour tout $n > 0$ par :

$$b_n = 0 \quad \text{et} \quad a_n = F^{-1}\left(\frac{1}{n}\right)$$

Définition 1.5.3. La fonction de quantile extrême dans le domaine d'attraction de Fréchet. est défini par :

$$q(\alpha) = \alpha^{-\xi}l(\alpha^{-1}) \quad \text{avec } l \in RV_0. \quad (1.7)$$

avec RV : Variations régulières.

B. Domaine d'attraction de Weibull

Théorème 1.5.2. [35]. Une fonction de répartition F appartient au domaine d'attraction Ψ_α (de Weibull avec un indice des valeurs extrêmes $\xi < 0$) si et seulement si $x_F < 1$ et si la fonction de répartition F^* définie par :

$$F^*(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ F(x_F - \frac{1}{x}) & \text{si } x \geq 0, \end{cases}$$

Une fonction de répartition F du domaine d'attraction de Weibull s'écrit pour $x \leq x_F$:

$$F(x) = 1 - (x_F - x)^{-1/\xi}l((x_F - x)^{-1}), \quad l \in RV_0$$

Dans ce cas un choix possible pour les suites a_n et b_n est :

$$b_n = x_F \quad \text{et} \quad a_n = x_F - F^{-1}\left(1 - \frac{1}{n}\right)$$

Définition 1.5.4. le quantile extrême dans le domaine d'attraction de Weibull d'ordre $\alpha \in [0, 1]$ associé s'écrit :

$$q(\alpha) = x_F - \alpha^{-\xi}l(1/\alpha) \quad \text{avec } l \in RV_0. \quad (1.8)$$

C. *Domaine d'attraction de Gumbel*

Théorème 1.5.3. [35]. *Une fonction de répartition F appartient au domaine d'attraction de Λ (Gumbel avec un indice des valeurs extrêmes $\xi = 0$) si et seulement si il existe $z < x_F \leq \infty$ tel que :*

$$\bar{F}(x) = c(x) \exp \left\{ - \int_z^x \frac{1}{a(t)} dt \right\} \quad \text{avec } z < x < x_F.$$

Où $c(x) \rightarrow c > 0$ lorsque $x \rightarrow x_F$ et a est une fonction positive et dérivable de dérivée a' telle que $a'(x) \rightarrow 0$ lorsque $x \rightarrow x_F$.

Dans ce cas, un choix possible pour les suites (a_n) et (b_n) pour tout $n > 0$ est :

$$a_n = q(1/n) \quad \text{et} \quad b_n = \frac{1}{\bar{F}(a_n)} \int_{a_n}^{x_F} \bar{F}(z) dz$$

Remarque 1.5.1. Il est difficile de caractériser la fonction quantile de lois appartenant au domaine d'attraction de Gumbel. Du fait que ce domaine d'attraction regroupe une grande diversité de lois.

1.6 Conclusion

Dans ce chapitre, nous avons fait un petit rappel sur la statistique d'ordre et sur la théorie des valeurs extrêmes, en mentionnant les différentes caractéristiques et les notions de base qui sont indispensables pour le chapitre suivant où nous nous intéressons aux méthodes d'estimations de quantiles.

Chapitre 2

Estimation de quantiles

2.1 Introduction

L'estimation de quantiles ou l'estimation de la fonction quantile a été intensivement étudiée par Azzalin (1981)[3], Chen et Tang (2005)[6]. Nayadara (1964)[28], Parzen (1979)[30] et McNeil et al (2005)[27]. Mais la plupart de ces estimateurs souffrent soit d'un biais ou d'inefficacité pour les niveaux de probabilité élevée.

Il existe en théorie des valeurs extrêmes différentes approches pour estimer la fonction quantile, nous considérons dans ce chapitre les méthodes suivantes :

- La méthode paramétrique.
- La méthode semi-paramétrique.
- La méthode des excès.
- La méthode non-paramétrique.

2.2 Approche paramétrique

Estimateur du maximum de vraisemblance

Cette méthode notée *E.M.V* est fréquemment utilisée en statistique. L'estimation par le maximum de vraisemblance donne des résultats asymptotiques efficaces, et les estimateurs obtenus convergent sous certaines conditions vers les vraies valeurs des paramètres. Soit (X_1, \dots, X_n) un échantillon de n variable aléatoire au point $x = x_1, \dots, x_n$, supposées indépendantes identiquement distribuées de densité h_θ , où $\theta = (\mu, \sigma, \xi)$. L'expression de la fonction de vraisemblance donnée par :

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n h_\theta(x_i).$$

l'estimateur $\hat{\theta}$ est donné par la résolution du système suivant :

$$\begin{cases} \frac{\partial \log L}{\partial \theta} = 0, \\ \frac{\partial^2 \log L}{\partial^2(\theta)} < 0. \end{cases}$$

Estimation paramétrique de quantiles

Supposons que F appartient à une famille de distributions paramétriques $F = \{F_\theta, \theta \in \Theta \subset \mathbb{R}^k\}$. L'idée d'estimation paramétrique est de supposer que toute quantité statistique peut être vue en fonction de θ et donc un estimateur naturel est obtenu en substituant un estimateur de paramètre $\hat{\theta}$ pour θ , ensuite étant donné que

$$q(\alpha) = F_\theta^{-1}(\alpha),$$

donc un estimateur naturel est :

$$\hat{q}(\alpha) = F_{\hat{\theta}}^{-1}(\alpha). \quad (2.1)$$

Avec le paramètre θ est estimé par la méthode de vraisemblance.

2.3 Approche des quantiles extrêmes par la loi des valeurs extrêmes

Principe de la méthode

Soit X_1, \dots, X_n n-échantillon de variable aléatoire i.i.d de densité de probabilité f et de fonction de répartition F . Pour estimer le quantile q_n , Guida et Longo (1988) [20] utilisent l'approximation

$$\mathbb{P}\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \approx H_\xi(x). \quad (2.2)$$

où a_n et b_n sont des suites de normalisation et ξ c'est l'ndice de valeur extrême.

Où de façon équivalente :

$$\mathbb{P}(X_{n,n} \leq z) \approx H_\xi\left(\frac{z - b_n}{a_n}\right). \quad (2.3)$$

Dans la suite on introduit :

$$H_{\xi, a_n, b_n} = H_\xi\left(\frac{z - b_n}{a_n}\right).$$

L'expression (2.3) montre que la loi du maximum $X_{n,n}$ peut être approchée par H_{ξ, a_n, b_n} .

On peut réécrire l'équation (2.2) comme suit :

$$\lim_{n \rightarrow \infty} n \log(F(a_n x + b_n)) = \lim_{n \rightarrow \infty} n \log(1 - \bar{F}(a_n x + b_n)) = \log(H_\xi(x)).$$

Comme $n \rightarrow \infty$, on peut montrer que $a_n x + b_n \rightarrow x_F$ et par conséquent $\overline{F}(a_n x + b_n)$ converge vers 0.

Un développement limité au premier ordre de $\log(1 + u)$ donne donc :

$$\overline{F}(a_n x + b_n) \approx -\frac{1}{n} \log(H_\xi(x)).$$

Pour tout ξ , on peut alors approcher le quantile q_α par :

$$q(\alpha) \approx b_n + a_n x_\alpha \quad (2.4)$$

où x_α vérifie $-\log(H_\xi(x_\alpha)) = n\alpha$.

On a alors un estimateur de quantile extrême de type suivant :

$$\hat{q}(\alpha) \approx \hat{b}_n + \hat{a}_n x_\alpha = \begin{cases} \hat{a}_n (n\alpha)^{-\hat{\xi}} + \hat{b}_n, & \text{si } F \in D(\text{Fréchet}) \\ -\hat{a}_n (n\alpha)^{-\hat{\xi}} + \hat{b}_n, & \text{si } F \in D(\text{Weibul}) \\ -\hat{a}_n \log(n\alpha) + \hat{b}_n, & \text{si } F \in D(\text{Gumbel}) \end{cases} \quad (2.5)$$

où (\hat{a}_n, \hat{b}_n) et $\hat{\xi}$ sont respectivement des estimateurs des suites (a_n, b_n) et de l'indice de queue ξ . Dans le cas particulier où $\xi = 0$, les auteurs proposent d'utiliser l'approche basée sur la loi *GEV*.

Définition 2.3.1. [12]. L'estimateur du quantile extrême de la loi *GEV* est défini par :

$$\hat{q}_n^{gev}(\alpha) = \hat{b}_n + \frac{\hat{a}_n}{\hat{\xi}_n} \left(\left(\frac{1}{n\alpha} \right)^{\hat{\xi}_n} - 1 \right).$$

Théorème 2.3.1. (*Weinstein (1973)*) [44]. Soit $F \in D(\text{Gumbel})$, il existe deux suites $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ telles que pour tout $x \in \mathbb{R}$ et $\nu > 0$,

$$\lim_{n \rightarrow \infty} \overline{F} \left[(b_n^\nu + c_n x)^\frac{1}{\nu} \right] = \exp(-x), \quad (2.6)$$

où, $c_n = a_n \nu b_n^{\nu-1}$.

Dans une telle situation, on approche le quantile par

$$q_{\alpha_n} \approx (b_n^\nu + c_n x_{\alpha_n})^\frac{1}{\nu} \text{ où } x_{\alpha_n} \text{ vérifie } \exp(-x_{\alpha_n}) = n\alpha_n,$$

et un estimateur du quantile extrême est obtenu en remplaçant les suites b_n et c_n respectivement par leurs estimateurs \hat{b}_n et \hat{c}_n , i.e

$$\hat{q}_{\alpha_n} \approx \left(\hat{b}_n^\nu - \hat{c}_n \log(n\alpha_n) \right)^\frac{1}{\nu}. \quad (2.7)$$

Les paramètres ξ , a_n , b_n et c_n de ces lois peuvent être estimés par la méthode du maximum de vraisemblance (Prescott et Walden, 1980, 1983) [33] ou la méthode des moments (Hosking et al., 1985) [23]. Dans le cas de l'approche EVD, Smith (1985) [41] fait une étude

détaillée du comportement asymptotique des estimateurs des paramètres ξ , a_n , b_n et c_n obtenus par la méthode du maximum de vraisemblance. Toutefois, il est conseillé d'utiliser les estimateurs des moments pondérés car ceux-ci sont non seulement explicites et faciles à calculer mais aussi parce qu'ils donnent de meilleurs résultats que les estimateurs du maximum de vraisemblance quand on a des échantillons de petite ou de moyenne taille. La principale difficulté de l'estimation des paramètres ξ , a_n , b_n et c_n est due au fait qu'il faut un échantillon de maxima, lequel est parfois difficile à extraire des données initiales.

2.4 Approche des quantiles extrêmes par la méthode des excès P.O.T

2.4.1 Principe de la méthode

Cette méthode assure que la loi des observations qui dépassent un seuil u peut être approchée, pour u grand, par une loi de Pareto généralisée GPD. Le quantile extrême est alors estimé en inversant la fonction de répartition de la loi de Pareto généralisée et en estimant les paramètres de cette loi à l'aide des observations supérieures au seuil u .

Définition 2.4.1. [12]. L'estimateur du quantile extrême de la loi GPD est défini par :

$$\hat{q}_n^{gpd}(\alpha) = X_{n-k_n+1,n} + \frac{\hat{\sigma}_n}{\hat{\xi}_n} \left(\left(\frac{n}{k_n}(1-\alpha) \right)^{-\hat{\xi}_n} - 1 \right). \quad (2.8)$$

Où $\hat{\xi}$ et $\hat{\sigma}$ sont respectivement des estimateurs des paramètres de forme et d'échelle. k_n est le nombre d'excès (auparavant noté N_u) et $u = X_{n-k_n+1,n}$ pour k_n assez grand.

Remarque 2.4.1. On notera la similitude entre l'expression du quantile de la loi GEV et celle du quantile de la loi GPD. Il y a trois paramètres inconnus dans chacune d'entre elles :

- a. l'indice des valeurs extrêmes ξ qui est le même dans les deux expressions soulignant de fait son importance dans le comportement de la queue de distribution et donc celui des valeurs extrêmes.
- b. Le paramètre d'échelle σ joue le rôle de a_n dans l'approche GEV,
- c. Le seuil u joue le rôle de b_n dans l'approche GEV.

Pour pouvoir estimer des quantiles extrêmes, il nous faut donc estimer ces paramètres.

2.4.2 La detection du seuil

Le seuil u joue un rôle important pour cela le choix de ce dernier doit être suffisamment grand pour pouvoir utiliser les résultats asymptotique, mais pas trop élevé pour obtenir des estimateur précises ni pas trop faible pour ne pas prendre en considération des valeurs non

extrêmes. Dans ce sens, plusieurs méthodes de détection du seuil ont été proposées. La technique graphique la plus utilisée est la fonction d'excès en moyenne (mean excess plot ou mean residual life plot (ME-plot)).

Le Mean Excess Plot

Le Mean Excess Plot est le graphe des points $(u, e(u))$ où $e(u)$ est la moyenne des excès au déla du seuil u , définie par :

$$e(u) = E[X - u | X > u] \quad 0 < u < \infty$$

Elle correspond à une fonction de u qui s'exprime à l'aide de $1 - F$. Nous supposons que pour ce modèle proposé, $E(X) < \infty$.

Un estimateur empirique de cette fonction est donné par :

$$\hat{e}(u) = \frac{\sum_{i=1}^n x_i 1_{(u, \infty)}(x_i)}{\sum_{i=1}^n 1_{(u, \infty)}(x_i)} - u$$

telle que $1_{(u, \infty)}$ est égale à 1 si $x_i > u$ et 0 sinon, et nous choisissons la $(k + 1)$ ème grande observation : $u = x_{n-k_n+1, n}$ comme seuil, d'où l'estimation

$$\hat{e}(x_{n-k_n+1, n}) = \frac{1}{k} \sum_{j=1}^k x_{n-j+1, n} - x_{n-k_n+1, n}.$$

Proposition 2.4.1. [12]. *Si X_1, \dots, X_{N_u} suivent une loi $GPD_{\xi, \sigma}$, alors pour $\xi < 1$*

$$E(X - u | X > u) = \frac{\xi}{1 - \xi} u + \frac{\sigma}{1 - \xi}$$

Dans ce cas, le seuil à retenir u , est celui pour lequel la moyenne des excès est approximativement linéaire.

Plus les queues de la distribution sont épaisses, plus la fonction $e(u)$ tend rapidement vers l'infini.

En effet, la fonction moyenne des excès relative à une distribution de loi Pareto de paramètre $\alpha > 0$, appartenant au domaine d'attraction de Fréchet d'indice des valeurs extrêmes $\xi = \frac{1}{\alpha}$, s'écrit comme suit : $e(u) = \frac{k+u}{\alpha-1}$.

Alors que pour une distribution de queue moins épaisse, par exemple la distribution de Weibull de paramètres $\lambda > 0$ et $\nu > 0$, appartenant au domaine d'attraction de Gumbel, la fonction moyenne des excès relative à cette loi s'écrit comme suit : $e(u) = \frac{u^{1-\nu}}{\lambda\nu}$.

2.5 Approche semi-paramétrique

2.5.1 Lois à queues lourdes

On se restreint aux fonctions $F \in D(\text{Fréchet})$ pour lesquelles on a la caractérisation suivante

$$\bar{F}(x) = x^{-1/\xi}l(x),$$

Dans le domaine d'attraction de Fréchet, l'approche semi-paramétrique est basée sur la définition de la fonction quantile (1.7). Pour tout $\xi > 0$, on a :

$$q(\alpha_n) = \bar{F}^{\leftarrow}(\alpha_n) = \alpha_n^{-\xi}l(\alpha_n^{-1}) \quad \text{avec } l \in RV_0, \quad (2.9)$$

$$q(\beta_n) = \bar{F}^{\leftarrow}(\beta_n) = \beta_n^{-\xi}l(\beta_n^{-1}) \quad \text{avec } l \in RV_0. \quad (2.10)$$

En divisant (2.9) par(2.10) et à l'aide de la définition d'une fonction à variations lentes, si β_n est suffisamment petit et $\alpha_n < \beta_n$ on obtient l'approximation suivante :

$$q(\alpha_n) = \bar{F}^{\leftarrow}(\alpha_n) \simeq \bar{F}^{\leftarrow}(\beta_n) \left(\frac{\beta_n}{\alpha_n} \right)^{\xi}. \quad (2.11)$$

En remplaçant $\bar{F}^{\leftarrow}(\beta_n)$ et ξ par des estimateurs, on obtient l'estimateur de Weissman (1978) défini par [45] :

$$\hat{q}_{\alpha_n}^W = X_{n - \lfloor n\beta_n \rfloor + 1, n} \left(\frac{\beta_n}{\alpha_n} \right)^{\hat{\xi}}. \quad (2.12)$$

Pour les propriétés de l'estimateur de Weissman, on peut se référer à l'ouvrage de Embrechts et al.(1997) [13].

2.5.2 Estimation du paramètre de la loi des valeurs extrêmes

Dans la littérature de la théorie des valeurs extrêmes, on trouve plusieurs techniques semi-paramétriques pour l'estimation de l'indice de queue (ξ) qui joue un rôle essentiel dans le comportement de la loi des extrêmes.

Dans cette partie nous exposerons uniquement trois estimateurs de l'indice de queue sont : estimateurs Hill (1975), estimateur de Pickands (1975) et l'estimateur de Zipf.

A. Estimateur de Hill

Définition 2.5.1. [22] L'estimateur de Hill n'est utilisable que pour les distribution de Fréchet ($\xi > 0$).

Soit $(k_n)_{n \geq 1}$ une suite d'entiers avec $1 < k_n \leq n$, l'estimateur de Hill est défini par :

$$\hat{\xi}_{k,n}^H = \frac{1}{k_n - 1} \sum_{i=1}^{k_n-1} \log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n}).$$

La normalité asymptotique de cette estimateur a été démontrée par Smith (1987) [42].

B. *Estimateur de Pickands*

L'estimateur de Pickands a l'avantage d'être valable quel que soit le domaine d'attraction de la distribution. Pickands (1975) démontre la consistance faible de son estimateur. La convergence forte ainsi que la normalité asymptotique ont été démontrées par Dekkers et de Haan (1989).

Définition 2.5.2. [32] On suppose que $\{X_i, i = 1, \dots, n\}$ est une suite de variables aléatoires indépendantes de loi F appartenant à l'un des domaines d'attractions. Soit $(k_n)_{n \geq 1}$ une suite d'entiers avec $1 \leq k_n < n$, l'estimateur de Pickands est défini par :

$$\hat{\xi}_{k_n}^P = \frac{1}{\log 2} \log \left(\frac{X_{n-k_n+1,n} - X_{n-2k_n+1,n}}{X_{n-2k_n+1,n} - X_{n-4k_n+1,n}} \right).$$

La convergence forte ainsi que la normalité asymptotique ont été démontrées par Dekkers et de Haan (1989) [9].

C. *Estimateur de Zipf*

En 1996, Kratz et Resnick [24] et Schultze et Steinebach [39] ont indépendamment proposé d'estimer l'indice de queue par la méthode des moindres carrés classique. Leur estimateur connu sous le nom de Zipf est asymptotiquement gaussien.

Définition 2.5.3. Soit $(k_n)_{n \geq 1}$ une suite d'entiers avec $1 < k_n \leq n$, l'estimateur de Zipf est défini par :

$$\hat{\xi}_{k_n}^Z = \frac{\frac{1}{k_n} \sum_{j=1}^{k_n} \log \frac{k_n+1}{j} \log X_{n-j+1,n} - \frac{1}{k_n} \sum_{j=1}^{k_n} \log \frac{k_n+1}{j} \left(\frac{1}{k_n} \sum_{j=1}^{k_n} \log X_{n-j+1,n} \right)}{\frac{1}{k_n} \sum_{j=1}^{k_n} \left(\log \frac{k_n+1}{j} \right)^2 - \left(\frac{1}{k_n} \sum_{j=1}^{k_n} \log \frac{k_n+1}{j} \right)^2}$$

Le choix du nombre k_n

La difficulté consiste à choisir le nombre k_n de statistiques d'ordre à utiliser. Les résultats concernant les estimateurs de l'indice des valeurs extrêmes sont asymptotiques : ils sont obtenus lorsque $k_n \rightarrow \infty$ et $k_n/n \rightarrow 0$. On doit donc s'assurer de ne pas trop s'éloigner de la queue de distribution là où se trouve l'information importante.

2.6 La méthode non paramétrique

Dans cette section on s'intéresse à l'estimation de quantile par la méthode du noyau.

2.6.1 Estimation non paramétrique de la densité de probabilité par la méthode du noyau

Un des plus vieux problèmes de la statistique non paramétrique consiste à estimer la densité de probabilité à partir d'un échantillon observé. Une des premières analyses consacrées à ce sujet est due au biométricien Karl Pearson, il y a une centaine d'années. Dans son article [31], une première approche (dite paramétrique) consiste à supposer que la densité de probabilité f appartient à une famille de densités qui peuvent être décrites par un petit nombre (connu) de paramètres réels. Ainsi, à moins d'avoir sur le phénomène aléatoire étudié des informations à priori très précises et indiscutables, le champ d'application d'un modèle paramétrique n'est satisfaisant que lorsque l'inflation du nombre de paramètres est telle que les méthodes d'estimation du modèle deviennent tout à fait inefficaces.

Pour pallier les insuffisances et les défauts des familles paramétriques, une seconde approche dite non paramétrique propose de "laisser parler les données", sans spécifier au préalable de forme sur f . Il existe plusieurs méthodes non paramétriques plus robustes que la méthode par histogramme : la méthode d'estimation par les séries orthogonales et la méthode du noyau. Nous regarderons brièvement en quoi consiste la méthode du noyau.

Principe de la méthode

On appelle estimateur à noyau k de f , l'estimateur donné par :

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right) \quad \text{pour} \quad \int_{-\infty}^{+\infty} k(u)du = 1 \quad (2.13)$$

où n est la taille de l'échantillon de X , k : la fonction noyau est associée à chaque observation de l'échantillon et h est le paramètre de lissage qui détermine son étendue.

k est une fonction de noyau définie sur \mathbb{R} .

Cependant, des contraintes ont été imposées sur ces deux paramètres, et certaines d'entre elles sont primordiales et intuitives :

(H1) Pour assurer la convergence de l'estimateur $\hat{f}_h(x)$, les seules conditions imposées sont :

$$h \rightarrow 0 \quad \text{et} \quad nh \rightarrow \infty \quad \text{quand} \quad n \rightarrow \infty$$

(H2) Soit le noyau k remplissant les conditions suivantes :

a. $\int_{-\infty}^{+\infty} k(x)dx = 1$ le noyau k est une densité de probabilité.

b. $\int_{-\infty}^{+\infty} xk(x)dx = 0$ le noyau k est un noyau symétrique.

c. $\int_{-\infty}^{+\infty} x^2k(x)dx = \sigma^2 \leq +\infty$

Noyaux usuels

Les noyaux les plus utilisés dans l'estimation de la densité de probabilité sont donnés dans le tableau suivant [47] :

<i>Noyau</i>	$\mathbf{k(u)}$
Uniforme	$\frac{1}{2}, \quad u \leq 1$
Triangulaire	$(1 - u), \quad u \leq 1$
Gaussien	$\frac{1}{2\pi} \exp(-\frac{u^2}{2}), \quad u \in \mathbb{R}$
Biweight	$\frac{15}{16}(1 - \mu^2)^2, \quad u \leq 1$
Triweight	$\frac{35}{32}(1 - \mu^2)^3, \quad u \leq 1$
Epanechnikov	$\frac{3}{4\sqrt{5}}(1 - \frac{\mu^2}{5}), \quad u \leq \sqrt{5}$
Cosine	$\frac{\pi}{4} \cos(\frac{\pi u}{2}), \quad u \leq 1$

TABLE 2.1 – Noyaux usuels.

2.6.2 Estimation de la fonction de répartition

La fonction de répartition joue un rôle important dans le domaine de la statistique. Si on connaît cette fonction, tous les paramètres importants (par exemple, la moyenne, la variance, les fractiles et les quantiles) s'ils existent deviennent connus. De ce fait, si on peut établir une estimation précise de la fonction de répartition d'une variable, on peut aussi définir des estimations précises des paramètres de la loi de cette fonction.

Il est vrai que l'on peut souvent passer d'un estimateur de la densité à un estimateur de la fonction de répartition par intégration et d'un estimateur de $F(x)$ à un estimateur de $f(x)$ par dérivation. Néanmoins une particularité est à souligner : c'est l'existence de la fonction de répartition empirique. La fonction de répartition empirique a de bonnes propriétés de convergence mais possède certains inconvénients comme celui de ne pas prendre en compte une éventuelle information supplémentaire ou bien le fait d'être une fonction en escalier. Dès que l'on restreint quelque peu le modèle envisagé pour les données, il existe des estimateurs qui sont préférables à la fonction de répartition empirique par exemple, l'estimateur à noyau.

Définition 2.6.1. Une estimation de la fonction de distribution F peut être obtenue par

une simple intégration de la fonction f_h . Ainsi, l'estimateur à noyau de F est donné par

$$\begin{aligned}\hat{F}_h(x) &= \int_{-\infty}^x \hat{f}_h(t) dt \\ &= \frac{1}{n} \sum_{i=1}^n k\left(\frac{x-x_i}{h}\right), \quad k(x) = \int_{-\infty}^x k(t) dt.\end{aligned}$$

2.6.3 Choix du paramètre de lissage

Le paramètre de lissage h est un facteur important dans l'estimation par la méthode des noyaux. Il représente en quelque sorte une fenêtre qui centrée sur chaque observation, détermine le degré de lissage de l'estimation d'une fonction de densité. Un faible paramètre de lissage implique un faible degré de lissage et résulte en une fonction de densité irrégulière. Plusieurs méthodes pour choisir ce paramètre ont été proposées dans la littérature et quelques études comparatives intéressantes ont été effectuées sur ces méthodes.

1. Méthodes plug-in

La décision d'un choix optimal pour le paramètre de lissage h s'oppose la spécification d'un critère d'erreur qui puisse être optimiser. Dans ce cas on cherche à minimiser l'erreur quadratique moyenne intégrée asymptotique AMISE [29].

On a :

l'Erreur Quadratique Moyenne Intégrée MISE est de la forme :

$$MISE(f, \hat{f}_h) = \int \mathbb{E}(f(x) - \hat{f}_h(x))^2 dx = \frac{h^4}{4} \sigma_k^4 \int (f''(x))^2 dx + \frac{1}{nh} \int k^2(u) du + o\left(h^5 + \frac{1}{n}\right). \quad (2.14)$$

L'Erreur Quadratique Moyenne Intégrée Asymptotique est alors de la forme :

$$AMISE = MISE(f(x), \hat{f}_h(x)) - o\left(h^5 + \frac{1}{n}\right) = \frac{h^4}{4} \sigma_k^4 R(f'') + \frac{R(K)}{nh}. \quad (2.15)$$

Le paramètre de lissage h^* qui minimise l'Erreur Quadratique Moyenne Intégrée Asymptotique est de la forme :

$$h^* = \left[\frac{R(K)}{\sigma_K^4 R(f'')} \right]^{1/5} n^{-1/5}.$$

La valeur du AMISE optimale $AMISE^* = AMISE(h^*)$ est alors de forme

$$AMISE^* = \frac{5}{4} [\sigma_K R^4(K) R(f'')]^{1/5} n^{-4/5}.$$

Avec :

$$R(K) = \int k^2(u) du.$$

$$R(f'') = \int (f''(x))^2 dx.$$

σ^2 est la variance du noyau.

R^4 : désigne la dérivée quatrième du noyau.

2. Méthodes cross validation (validation croisée)

L'idée de base des méthodes validation croisée consiste à trouver une fonction de score $CV(h)$ ayant la même structure que le $MISE(h)$ et dont le calcul soit plus simple.

- **Validation croisée non biaisée**

Cette méthode a été proposée par Rudemo [36] en 1982 et Bowman [5] en 1984. Le critère consiste à choisir le paramètre de lissage qui minimise un estimateur convenable de :

$$ISE(h) = \int_{\mathbb{R}} [\hat{f}_h(x) - f(x)]^2 dx = \int_{\mathbb{R}} \hat{f}_h^2(x) dx - 2 \int_{\mathbb{R}} \hat{f}_h^2(x) f(x) dx + \int_{\mathbb{R}} f^2(x) dx.$$

Puisque $\int_{\mathbb{R}} f^2(x) dx$ ne dépend pas du paramètre de lissage h . On peut choisir le paramètre de lissage de façon à ce qu'il minimise un estimateur de :

$$UCV(h) = ISE(h) - \int_{\mathbb{R}} f^2(x) dx = \int_{\mathbb{R}} \hat{f}_h^2(x) dx - 2 \int_{\mathbb{R}} \hat{f}_h^2(x) f(x) dx.$$

On veut premièrement trouver un estimateur de $\int_{\mathbb{R}} \hat{f}_h^2(x) f(x) dx$. Remarquons que :

$$\int_{\mathbb{R}} \hat{f}_h^2(x) f(x) dx = \mathbb{E}[f_h(x)]$$

L'estimateur empirique de $\int_{\mathbb{R}} \hat{f}_h^2(x) f(x) dx$, est alors $\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i)$. Le critère à optimiser est alors :

$$UCV(h) = \int_{\mathbb{R}} \hat{f}_h^2(x) dx - \frac{2}{n} \sum_{i=1}^n f_{h,i}(x_i). \quad (2.16)$$

Où $f_{h,i}(x) = \frac{1}{(n-1)h} \sum_{1 \leq i \leq n, i \neq j} k\left(\frac{x-x_j}{h}\right)$ est l'estimateur de la densité construit à partir de l'ensemble de points sauf le point x_i .

En utilisant l'équation (2.16), le critère $UCV(h)$ devient :

$$UCV(h) = \frac{R(k)}{nh} + \sum_{i=1}^n \sum_{i \neq j, j=1}^n \left[\int \frac{1}{n^2 h^2} k\left(\frac{x-x_i}{h}\right) k\left(\frac{x-x_j}{h}\right) dx - \frac{2}{n(n-1)h} k\left(\frac{x_i-x_j}{h}\right) \right]. \quad (2.17)$$

avec $R(k) = \int k^2(u) du$

Nous noterons h_{ucv} l'estimateur de h qui minimise $UCV(h)$.

- **Validation croisée biaisée**

La méthode de validation croisée biaisée, a été introduit par Scott et Terrell [40] en 1987 pour remédier aux problèmes de validation croisée non biaisée. Il s'agit d'introduire un biais dans le UCV afin de réduire sa variance.

Lemme 2.6.1. (Scott et Terrell)

Supposant que le noyau k satisfait aux conditions suivantes :

$$\int k''(u)du = 0, \quad u_1(k'') = \int uk''(u)du = 0, \quad u_2(k'') = \int u^2k''(u)du = 2.$$

On obtient le développement asymptotique :

$$E[R(\hat{f}_h'')] = R(\hat{f}_h'') + \frac{R(k'')}{nh^5} + o(h^2).$$

avec :

$$R(\hat{f}_h'') = \int_{\mathbb{R}} (\hat{f}_h'')^2 dx \quad R(k'') = \int_{\mathbb{R}} (k''(u))^2 du$$

L'estimateur du *AMISE* :

$$BCV(h) = \frac{h^4}{4} \sigma_k^4 [R(f_h'') - \frac{R(k'')}{nh^5}] + \frac{R(k)}{nh}. \quad (2.18)$$

2.6.4 Estimation non paramétrique d'un quantile

Principe de la méthode

Soit X_1, X_2, \dots, X_n un échantillon d'une *v.a* X de fonction de répartition F et de densité f . $X_{(1)} \leq X_{(2)} \leq \dots X_{(n)}$ la statistique d'ordre correspondante. La fonction quantile q est définie comme suit :

$$q(\alpha) = \inf\{x : F(x) \geq \alpha\} \quad \text{avec } 0 < \alpha < 1$$

On note pour tout $0 < \alpha < 1$ le quantile d'ordre α de F par γ_α , $\gamma_\alpha = q(\alpha)$.

Définition 2.6.2. l'estimateur non paramétrique d'une fonction de répartition est la fonction empirique $F_n(x)$ sur la base de l'échantillon X_1, X_2, \dots, X_n est :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{X_i \leq x}$$

Où $1(A)$ est la fonction indicatrice de l'événement A donnée par :

$$1_A(x) = \begin{cases} 1, & \text{si } x \in A \\ 0, & \text{sinon} \end{cases}$$

l'estimateur non paramétrique de γ_α est le quantile empirique :

$$\hat{q}(\alpha) = \inf\{x : F_n(x) \geq \alpha\} = X_{([\!n\alpha\!] + 1)}$$

Où $[n\alpha]$ désigne la partie entière de $n\alpha$.

Inverse d'estimateur à noyau d'une fonction de répartition

Définition 2.1. L'estimateur à noyau d'un quantile le plus populaire est donné par Nadaraya (1964) [28]. Il est défini comme suit :

$$\hat{F}_h(x) = \frac{1}{n} \sum_{i=1}^n k_h(x - X_i)$$

où

$$k_h = \int_{-\infty}^x \frac{1}{h} k\left(\frac{t}{h}\right) dt,$$

k est la fonction noyau, l'estimateur du quantile $\hat{q}(\alpha) = \hat{F}_h^{-1}$ est défini par :

$$\hat{q}(\alpha) = \inf\{x : \hat{F}_h(x) \geq \alpha\}, \quad \text{avec } 0 < \alpha < 1. \quad (2.19)$$

Définition 2.2. Un type d'estimateur à noyau d'un quantile est donné par Yang(1985)[46] et également par Parzen(1979) [30]. Il est défini par :

$$\hat{q}(\alpha) = \sum_{i=1}^n X_{(i)} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h} K\left(\frac{\alpha - x}{h}\right) dx = \sum_{i=1}^n \left[K\left(\frac{i/n - \alpha}{h}\right) - K\left(\frac{i-1/n - \alpha}{h}\right) \right] X_{(i)} \quad (2.20)$$

Cette estimateur est une adaptation de la régression lissée proposé par Nadaraya (1964)[28] et Watson (1964)[43].

2.7 Conclusion

Nous avons présenté dans ce chapitre les différentes méthodes d'estimation de la fonction quantile. Nous avons présenté quelques cas particulier d'estimateurs basé sur l'approche paramétrique et ceux associés aux approches semi-paramétrique et non paramétrique.

Chapitre 3

Valeur à Risque et estimation

3.1 Introduction

La notion de Value-at-Risk (VaR) est apparue pour la première fois dans le secteur de l'assurance. A la fin des années 1980, la banque Bankers Trust fut l'une des premières institutions à utiliser cette notion sur les marchés financiers aux Etats-Unis, mais c'est principalement la banque JP Morgan qui dans les années 90 a popularisé ce concept notamment grâce à son système RiskMetrics (pour un historique complet de la notion de Value-at-Risk et de sa diffusion se reporter au livre de Dowd, 2005)[11]. La Value-at-Risk est ensuite devenue, en moins d'une dizaine d'années, une mesure de référence du risque sur les marchés financiers. De façon générale, la Value-at-Risk est définie comme la perte maximale de potentielle qui ne devrait être atteinte qu'avec une probabilité donnée sur un horizon temporel donné (Engle et Manganelli, 2001)[26]. La Value at Risk est donc la pire perte attendue sur un horizon de temps donné pour un niveau de confiance donné. Cette définition très simple constitue l'un des principaux attraits de la Value-at-Risk : il est en effet très facile de communiquer sur la VaR et de ainsi proposer une mesure homogène et générale de l'exposition au risque. Ainsi, la Value-at-Risk n'est rien d'autre qu'un fractile de la distribution de perte et profit associée à la détention d'un actif ou d'un portefeuille d'actifs sur une période donnée. La mesure de Value-at-Risk ne fait que refléter l'information contenue dans la queue gauche de la distribution des rendements d'un actif.

3.2 Les mesures de risque

Définition 3.2.1. On appelle mesure de risque toute application ρ associant un risque X un réel $\rho(X) \in \mathbb{R}$.

Après avoir défini ce qu'est une mesure de risque, nous rappelons les principales propriétés qu'elles doivent vérifier pour être jugées satisfaisantes.

3.2.1 Caractéristiques d'une mesure de risque

La définition d'une mesure de risque est très générale puisque toute fonctionnelle réelle positive d'une variable aléatoire peut être considérée comme étant une mesure de risque. Aussi, en pratique, on exige pour de telles mesures qu'elles disposent de propriétés mathématiques dont la transcription conceptuelle permette de les juger. En pratique, on exige fréquemment qu'une mesure de risque ρ qui est définie comme suit :

$$\rho : V \rightarrow \mathbb{R}$$

avec V l'ensemble des variables aléatoires à valeurs réelles.

cette mesure de risque possède une partie des caractéristique suivantes :

- **Homogénéité positive (respect des échelles) :**
 $\rho(\lambda x) = \lambda \rho(x)$ pour toutes variables x et tous les nombres réels positifs λ
- **Sous-additivité :**
 $\rho(x + y) \leq \rho(x) + \rho(y)$ pour toutes variables x et y . Notons qu'il est prouvé qu'une fonction homogène positive est convexe seulement si elle est sous-additive.
- **Monotonie :**
 $x \leq y$ implique que $\rho(x) \leq \rho(y)$ pour toutes variables x et y .
- **Invariance transitionnelle :**
 $\rho(x + \alpha r_0) = \rho(x) - \alpha$ pour toute variable x et nombres réels α , ainsi que pour le paramètre r_0 que nous pouvons interpréter comme le taux sans risque.

3.2.2 Mesure de risque cohérente

Une mesure de risque invariante par translation, sous-additive, homogène et monotone est dite cohérente.

3.2.3 Mesure de risque monétaire

Une mesure invariante par translation et monotone est dite monétaire.

3.3 Définitions

- **Risque de marché**
 Le risque de marché peut se définir comme le risque qu'un changement dans les prix des titres (actions) ou des produits dérivés dans lesquels la banque détient une position cause une perte. Ou bien, le risque de marché est le risque de perte qui peut résulter des fluctuations des prix des instruments financiers qui composent un

portefeuille.

- **Portefeuille**

Le portefeuille boursier, comme son nom l'indique, représente l'ensemble des titres sur lesquels un investisseur a investi sur le marché financier.

- **Action**

Une action est tout simplement une part de la propriété d'une entreprise. L'action représente un droit sur l'actif et le bénéfice. Plus vous achetez d'actions, plus l'intérêt que vous détenez dans l'entreprise est grand.

- **Indice boursier**

Un indice boursier constitue un indicateur clé pour déterminer la performance d'un marché. Les indices boursiers permettent aux investisseurs de gérer et analyser leur portefeuille d'actions.

- **Actif**

Tous objets, titres, propriétés ayant une valeur monétaire.

- **Actif financier**

Un actif financier est un titre ou un contrat, généralement transmissible et négociable (par exemple sur un marché financier), qui est susceptible de produire à son détenteur des revenus ou un gain en capital, en contrepartie d'une certaine prise de risque.

- **Marché financier**

Un marché financier est un marché sur lequel des personnes, des sociétés privées et des institutions publiques peuvent négocier des titres financiers, matières premières et autres actifs, à des prix qui reflètent l'offre et la demande. Les titres comprennent des actions et des obligations, ainsi que des produits de base, notamment des métaux précieux ou des produits de base agricoles.

3.4 Valeur à Risque (Value at risk) VaR

Définition 3.4.1. la VaR est une mesure statistique synthétique de la perte maximale qui peut être attendue sur une position dans des conditions de marché normales. C'est une méthode qui permet de communiquer le risque agrégé de perte possible d'un portefeuille en un simple chiffre à des non spécialistes. Par définition, la VaR est le niveau de perte maximal qui ne sera dépassé qu'avec une probabilité bien déterminée sur une période de temps. Statistiquement, elle est le quantile d'ordre $1 - \alpha$ de la distribution de perte et profit du portefeuille considéré.

la VaR dépend de deux données essentielles : le niveau de confiance et l'horizon de temps :

1. Taux de couverture et Niveau de Confiance

Le niveau de confiance choisi est un paramètre compris entre 0 et 1 (95% ou 99% en général) qui permet de contrôler la probabilité que l'on obtienne un rendement supérieur ou égale à la Value-at-Risk.

2. Horizon de détention

Le deuxième élément fondamental dans le calcul de la Value-at-Risk est la période de détention de l'actif ou du portefeuille d'actifs. La formule de calcul de la Value-at-Risk doit alors être ajustée de façon à tenir compte de la composition des rendements.

Qu'est-ce que la VaR représente concrètement ?

Elle représente en fait le montant de pertes, exprimé normalement en rendements (r_i), sur un horizon temporel h qui ne devrait généralement pas excéder un certain niveau de confiance $(1 - \alpha)$ donné. Cette mesure statistique doit ainsi refléter les pertes dues au risque de marché provenant d'une variation normale.

L'équation générique de la Valeur à Risque peut être alors posée de la façon suivante :

$$P(L_t \leq VaR) = \alpha, \quad (3.1)$$

où L_t représente la variable aléatoire perte. La VaR n'est alors que le quantile q_α de la variable perte L_t . L'équation (3.1) montre que la connaissance de la fonction de répartition est importante afin de déterminer la VaR. En effet :

$$VaR_\alpha = F_{L_t}^{-1}(\alpha).$$

Où $F_{L_t}(\cdot)$ est la fonction de répartition de la variable aléatoire L_t et $F_{L_t}^{-1}(\cdot)$ la fonction inverse associée, nommée aussi le quantile de L_t .

3.4.1 Interprétation mathématique

Selon Esch, Kieffer et Lopez (1997) et Jorion (2000), la VaR de l'actif en considération, pour une durée t et un niveau de probabilité α , se définit comme le montant de la perte étendue de sorte que ce montant, pendant la période $[0, t]$, ne devrait pas être plus important que la VaR avec une probabilité de $(1 - \alpha)$. Autrement dit :

$$P[L_t > VaR_\alpha] = 1 - \alpha \Leftrightarrow P[L_t < VaR_\alpha] = \alpha.$$

En centrant et réduisant l'expression, nous obtenons :

$$P \left[\frac{L_t - E(L_t)}{\sigma(L_t)} \leq \frac{VaR_\alpha - E(L_t)}{\sigma(L_t)} \right] = \alpha.$$

Nous pouvons donc définir :

$$\frac{VaR_\alpha - E(L_t)}{\sigma(L_t)} = Z_\alpha.$$

D'où nous pouvons calculer la VaR comme étant :

$$VaR_\alpha = E(L_t) + Z_\alpha \sigma(L_t).$$

3.4.2 Propriétés de la VaR

Lemme 3.4.1. Pour tout $\alpha \in [0, 1]$, si g est une fonction strictement croissante continue à gauche,

$$VaR_\alpha(g(X)) = F_{g(X)}^{-1}(\alpha) = g(F_X^{-1}(\alpha)) = g(VaR_\alpha(X)),$$

et si g est une fonction strictement décroissante, continue à droite et F_X bijective, on a

$$VaR_\alpha(g(X)) = F_{g(X)}^{-1}(\alpha) = g(F_X^{-1}(1 - \alpha)) = g(VaR_{(1-\alpha)}(X)).$$

3.5 Estimation de la VaR

Il existe plusieurs méthodes pour estimer la VaR comme :

- La méthode paramétrique ou la méthode de la matrice des variances-covariances.
- La méthode de simulation historique.
- La méthodes semi paramétriques (Valeurs extrêmes).

3.5.1 La méthode paramétrique

La VaR paramétrique est une évaluation analytique du portefeuille d'instruments financiers. La méthode paramétrique la plus reconnue provient de la banque JP Morgan qui développa ce modèle en 1994 et connu sous le nom de Risk Metrics [11]. C'est un modèle statistique qui suppose que le rendement ou bien le changement de prix d'un portefeuille suit, par exemple, souvent une loi normale. Une conséquence de cette décision est que cette même distribution est limitée par une matrice de variance covariance qui représente les corrélations entre ces facteurs. Suite aux hypothèses qui sont posées, il devient possible d'inférer une formule analytique du calcul de la VaR.

De manière générale, il est possible d'écrire alors la VaR de la manière suivante [11] :

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

Où μ représente le rendement espéré du portefeuille et σ c'est l'écart type de portefeuille. $\Phi^{-1}(1 - \alpha)$ est la fonction quantile de la distribution normale centrée réduite.

3.5.2 La méthode historique

Cette méthode est la plus intuitive et la plus simple à utiliser. Elle suppose une stationnarité des rendements, ce qui signifie que la distribution des rendements observée à partir des données historiques se reproduira à l'avenir. Cette approche utilise des données de marché plus ou moins récentes pour déterminer la distribution des variations des facteurs de risque qui peuvent simplement être les prix des actifs dans le cas des actions [37]. Ces variations historiques sont ensuite appliquées au portefeuille actuel dont on calcule la valeur pour chaque scénario historique. La distribution des valeurs possibles du portefeuille est ainsi estimée puis on extrait le quantile correspondant à un niveau de probabilité choisi. L'élément clé de cette approche reste donc la sélection de l'historique utilisé pour les calculs. Les résultats seront différents selon la fenêtre historique retenue et si l'on souhaite par exemple donner plus d'importance aux données de marché récentes ou à des périodes "agitées" (scénarii extrêmes). La simulation historique reste la méthode la plus couramment utilisée car elle est la plus facile à mettre en place malgré la nécessité de maintenir des historiques nombreux et longs selon le nombre d'actifs pris en compte.

3.5.3 La méthode des excès

La VaR n'est rien d'autre qu'un quantile extrême calculé à partir de la loi asymptotique des extrêmes (Distribution de Pareto Généralisée GPD), obtenue en modélisant les pertes (ou profits) extrêmes par la méthode des excès. Ainsi, la VaR_α correspondant au modèle GPD inconditionnel, pour un horizon donné h et une certaine probabilité α , ou de façon équivalente à un niveau de confiance $1 - \alpha = 0.95$ est :

$$VaR_\alpha = \mu + \frac{\hat{\sigma}}{\hat{\xi}} \left(\left(\frac{n}{N_\mu} (1 - \alpha) \right)^{-\xi} - 1 \right).$$

où $\hat{\xi}$ et $\hat{\sigma}$ représentent les estimateurs des paramètres de la loi GPD.

3.6 Conclusion

Dans ce chapitre, nous avons présenté une revue de littérature portant sur la définition de la VaR ainsi que les différentes étapes associées à son calcul et à la fin on a cité les techniques d'estimation de ce dernier.

Chapitre 4

Application numérique

4.1 Introduction

Durant ces dernières années, la Valeur à Risque est devenue un indicateur de risque largement utilisé par les établissements financiers, elle permet en effet d'appréhender les risques de marchés de façon globale dans une unité de mesure commune, quelle que soit la nature des risques (taux, change, actions, ... etc.).

Les baisses de prix peuvent affecter la position en fonds propres de la banque. Pour cette raison, les banques surveillent leur portefeuille de marché à l'aide d'outils tels que la VaR. Cependant, la VaR est un outil utilisé pour mesurer et quantifier le risque de marché pour une banque. Le risque de marché se réfère au portefeuille d'instruments négociables que la banque détient. La VaR met en évidence le pire des cas pour les prochaines 24 heures. Elle est le montant maximum que la banque risque de perdre en cas de problème ou de crise sur les marchés financiers. La VaR est tournée vers l'avenir, c'est une estimation de la taille des pertes futures. Comme nous ne pouvons être certains de ces évolutions futures du marché, il est courant de définir un niveau de confiance par exemple 95%. cela signifie simplement que 95% des cas, la perte ne devrait pas dépasser cette mesure de risque.

La question principale posée dans cette partie est : comment calculer la valeur à risque suivant les différentes méthodes ?

4.2 Type des données

Dans cette étude on s'intéresse au prix de clôture (voir l'annexe figure 4.9 page 58). Tel que le cours de clôture est le dernier prix enregistré à la fermeture du marché pour la journée. L'obtention de ce prix n'est pas certaine au cas où vous achetez le titre le lendemain, car le cours évolue constamment (même après la fermeture de la bourse pour la journée). Le cours de clôture est seulement un indicateur du rendement antérieur et, sauf en des circonstances extrêmes, il constitue une approximation de ce que vous devriez payer.

Pour mener à bien notre étude, on a extrait nos données du site web WWW.Yahoo.Finance.com. Prenons l'exemple de l'action Koninklijke Volker Wessels (KVV) qui cotée à la bourse d'Amsterdam avec des prix en Euros. La période qu'on a choisi est celle entre 17 avril 2018 et 1 avril 2020. On dispose en tout de 501 observations correspondantes aux prix de clôture journalier de l'action KvW.

L'action Koninklijke Volker Wessels

Koninklijke VolkerWessels est un groupe de construction côté aux activités diversifiées pratiquant le mot d'ordre. kvW opère principalement aux Pays-Bas, au Royaume-Uni, en Amérique du Nord et en Allemagne. Sur le plan opérationnel, son activité est organisée en six segments. Dans les pays où VolkerWessels opère, le groupe compte plus de 120 entreprises d'exploitation locales, possédant des bureaux et des directions au niveau national et régional.

La réalisation de cette application nécessite de faire appel à des commandes et des fonctions prédéfinies sous le logiciel R.

Le logiciel R : est un logiciel de statistique créé par Ross Ihaka et Robert Gentleman. Il est à la fois un langage informatique et un environnement de travail : les commandes sont exécutées grâce à des instructions codées dans un langage relativement simple, les résultats sont affichés sous forme de texte et les graphiques sont visualisées directement dans une fenêtre propre. Ce logiciel sert à manipuler des données, à tracer des graphiques et à faire des analyses statistiques sur ces données. Le logiciel *R* constitue aujourd'hui un langage de programmation intégré d'analyse statistique [21].

4.3 Analyse descriptive des prix de clôture

Le résumé statistique des données du tableau (4.9) est présenté dans le tableau suivant :

Min.	1st Qu.	Median.	Mean.	3rd Qu.	Max.
12.24	16.27	17.59	18.01	20.79	22.15

TABLE 4.1 – Description de la série des prix de clôture.

Le graphe des prix de clôture

La représentation graphique de l'échantillon étudié sur les deux dernières années est donnée dans la figure (4.1).

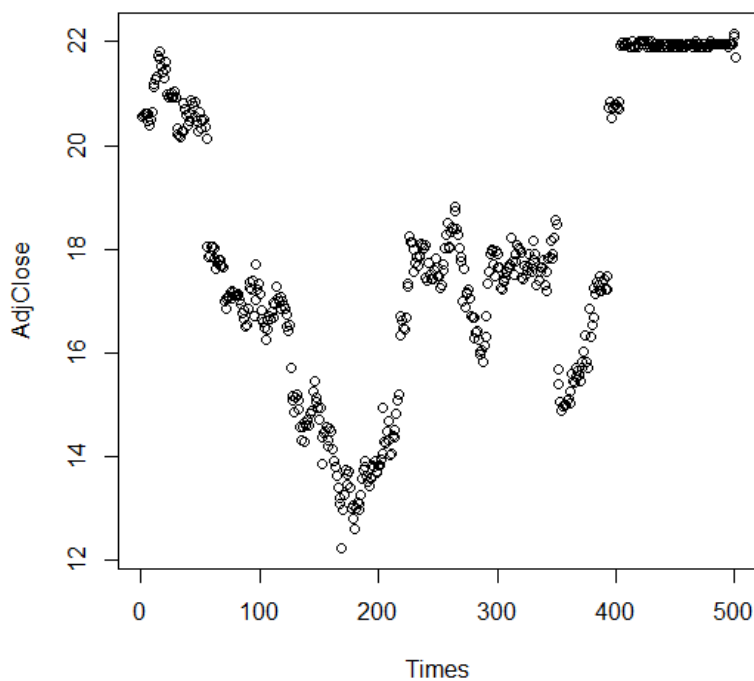


FIGURE 4.1 – Graphique des prix de clôture entre Avril 2018 - Avril 2020.

Intrépritation

D'après le graphe (4.1), on constate que le prix de clôture n'est pas stable. A la 200-ème journée, le prix de clôture est au plus bas, par contre a 400 jours, le prix de clôture devient stable.

Histogramme des prix de clôture

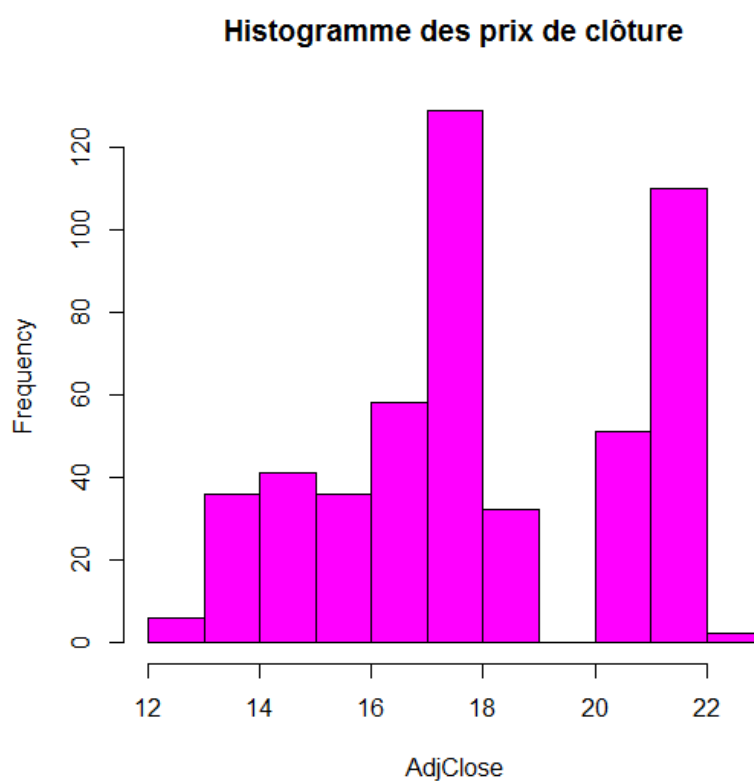


FIGURE 4.2 – l'histogramme des prix de clôture

Sur l'axe verticale du graphe (4.2), on trouve le nombre de fois qu'un prix de clôture s'est produit. Ceci est la fréquence.

4.4 Calcul des rendements journaliers

Le rendement périodique journalier pour la période i est calculé par la formule des rendements logarithmiques [15] :

$$r_i = \ln \left(\frac{p_i}{p_{i-1}} \right)$$

Avec : p_i lprix des actifs à la fin de la période i .

La commande utilisé sous R pour le calcul des redements est la suivante :

```
> Deltaprix = diff(log(AdjClose))
```

```
> Deltaprix
```

```
[1] 0.0008752554 0.0017481189 0.0000000000 0.0000000000 -0.0070112283
[6] -0.0044074865 0.0061649684 0.0069990569 0.0258190753 -0.0025522429
[11] 0.0067911986 0.0016906295 0.0185952155 0.0042752698 -0.0059905825
. . . . .
[491] 0.0000000000 0.0000000000 0.0000000000 -0.0022805483 0.0022805483
[496] 0.0000000000 0.0000000000 0.0068104234 0.0022598880 -0.0205251899
```

Le résumé statistique de ces données est présenté dans le tableau suivant :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.1645321	-0.0062615	0.0000000	0.0001078	0.0064773	0.1709578

TABLE 4.2 – Description de la serie des variations des prix.

Le graphe des logarithmes des rendements est donné dans la figure (4.3) :

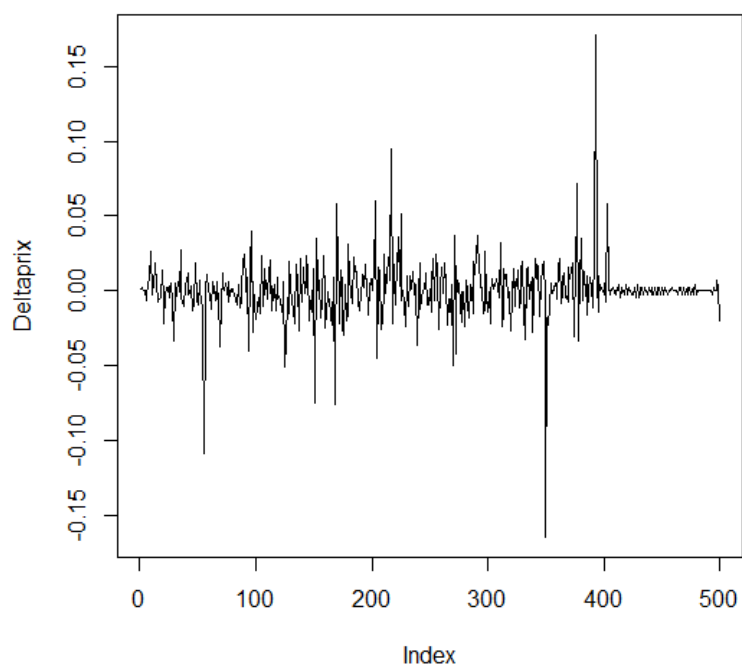


FIGURE 4.3 – Rendements logarithmiques journaliers du KVV.

Intrprétation

On constate d'après le graphe (4.3) que le rendement se fluctue autour de la moyenne, il a une tendance normale. On peut montrer cette remarque avec le test de normalité.

Histogramme des rendements

A partir des rendements de la série calculés sur la période allant du 17 Avril 2018 au 1 avril 2020, nous avons tracé l'histogramme des variations de prix.

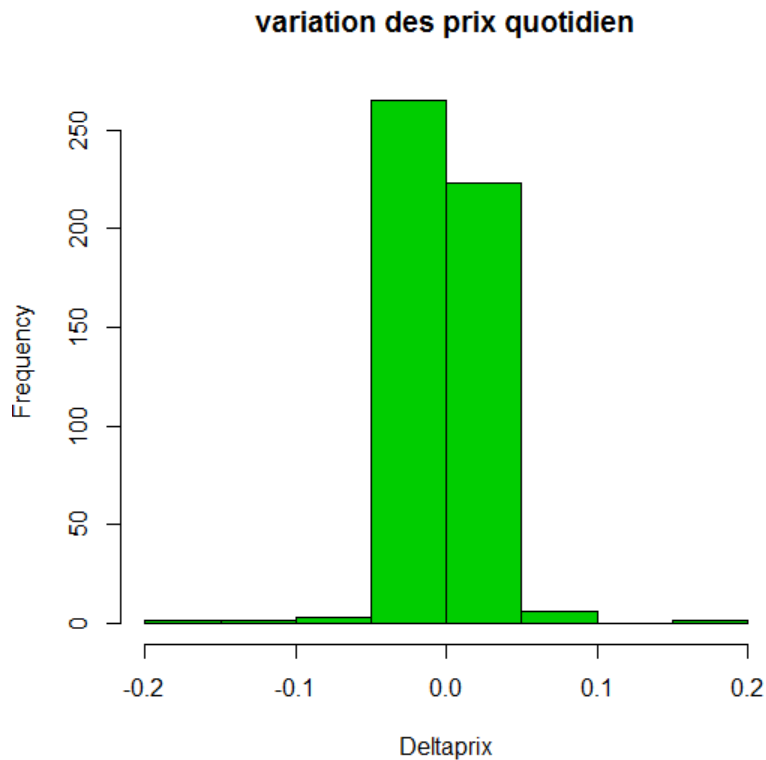


FIGURE 4.4 – Histogramme des variations des prix

Interprétation

Sur l'axe verticale du graphe (4.4), on trouve le nombre de fois qu'un certain changement de prix s'est produit. Ceci est la fréquence.

D'après le graphe (4.4), nous constatons que la plupart des jours, le changement de prix est nul ou proche de zéro.

Etude de la normalité

L'ajustement par la loi normale en utilisant le test de Shapiro Wilk

Le test de *Shapiro – Wilk* est un test permettant de savoir si une série de données suit une loi normale.

On teste les hypothèses suivantes :

H_0 : "L'échantillon suit une loi normale" contre H_1 : "L'échantillon ne suit pas la loi normale".

Où bien :

Hypothèse nulle : l'échantillon suit une loi normale. Par conséquent si la p -value du test est significative, l'échantillon ne suit pas une loi normale.

On fixe un seuil de signification $\alpha = 0.05$

Le test de Shapiro-Wilk est disponible dans le package **stats**. La fonction correspondante est `shapiro.test`.

```
> shapiro.test(rnorm(500, mean(Deltaprix), sd(Deltaprix)))
```

Résultat de la commande :

Shapiro-Wilk normality test

```
data : rnorm(500, mean(Deltaprix), sd(Deltaprix))  
W = 0.99737, p-value = 0.6149
```

Décision du test :

L'ajustement des données de notre échantillon par la loi normal donne une p -value égale à 0.6149. Cette valeur est supérieure au seuil de signification. D'où elle renvoie une p -value non significative, donc l'échantillon suit une loi normale.

Histogramme de logs rendements

```
> hist(Deltaprix, prob = T, breaks = 15, col = "gray")
```

```
> curve(dnorm(x, mean(Deltaprix), sd(Deltaprix)), add = T, lwd = 2)
```

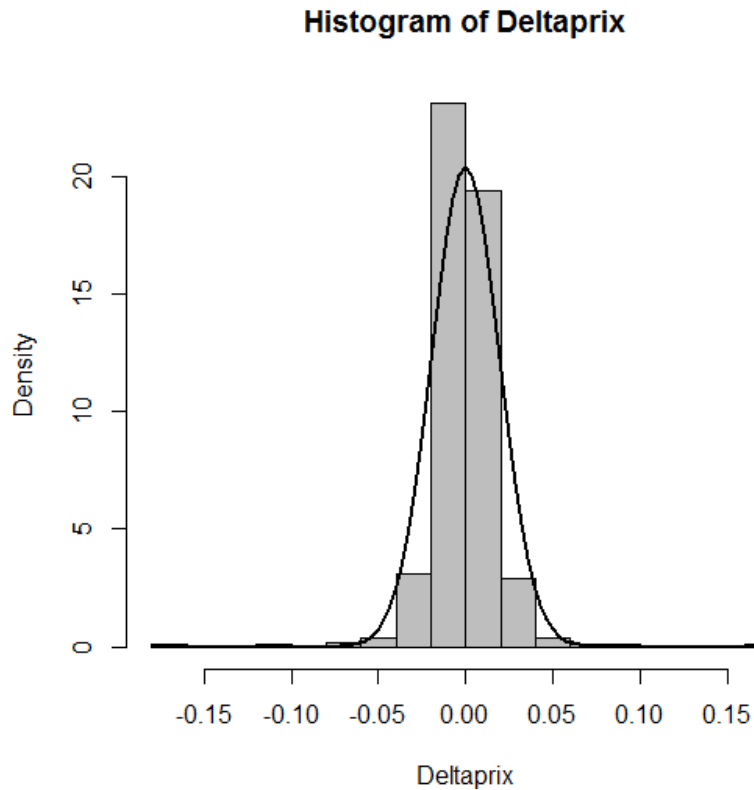


FIGURE 4.5 – Histogramme des variations des prix

4.5 La VaR par la méthode paramétrique

On rappelle que cette méthode suppose implicitement que le rendement de chaque actif est distribué suivant une loi normale de moyenne μ et de variance σ^2 . D'après le chapitre précédent (chapitre 3) nous avons :

$$VaR = \mu + Z_{(1-\alpha)}\sigma$$

Donc pour calculer la VaR des variations des prix, il faut calculer le quantile de la loi $N(0,1)$, et estimer μ et σ^2 par la moyenne et la variance empirique de la series des variations des prix.

Les résultats de la simulations sont donnés dans le tableau suivant :

$\Phi(0.95)$	$\hat{\mu}$	$\hat{\sigma}$	$VaR = \hat{\mu} + \Phi(0.95) * \text{sqrt}(\hat{\sigma})$
1.644542	0.03924841	0.9926054	1.677699

Cela signifie qu'il y a 95% de chances que la perte associée à la détention de l'actif n'excède pas 1.677699.

4.6 La VaR historique

La VaR Historique consiste à calculer la moyenne des quantiles de chaque mois pendant cette période.

Puisque la loi de distribution des prix de clôture est inconnue, on va estimer la *VaR* des variations des prix par un quantile empirique, avec un seuil $1 - \alpha = 0,95$.

```
> quantile(Deltaprix, 0.95)
```

```
[1] 95%  
0.02308707 % Quantile empirique au niveau 95%.
```

Pour calculer la *VaR* sur tout l'historique des prix de clôture, on utilise la démarche suivante

```
> N = length(Deltaprix)
> N
[1] 500

> varh = c()
> for(i in 31 : N)
+{
+varh[i - 30] = quantile(Deltaprix[(i - 30) : i], prob = 0.95, na.rm = TRUE)
+}
> Var_hist < -mean(varh)
> Var_hist

[1] 0.02223679 % VaR historique au niveau 95%.
```

Remarque 4.6.1. On peut calculer la VaR historique sur **EXCEL** en utilisant la fonction *CENTILE.EXCLURE* comme suit

$$CENTILE.EXCLURE(Deltaprix; 0.95) = 0,022326$$

$$Deltaprix = rendement = LN \left(\frac{p_i}{p_{i-1}} \right)$$

Avec $p_{i_1 \leq i \leq n}$ les prix de l'indice boursier à différentes dates i .

4.7 La VaR par la théorie des valeurs extrêmes

Dans cette partie on s'intéresse l'estimation de la VaR par la méthode POT. La VaR n'est rien d'autre que le quantile extrême calculé à partir de la loi asymptotique des extrema (Generalized Pareto Distribution), obtenu en modélisant les pertes (ou les profits) extrêmes par la méthode des excès.

4.7.1 Détection du seuil

La détermination du seuil est l'étape la plus délicate dans l'implémentation de l'approche *POT*, étant donné que la qualité du modèle en dépend. La convergence des excès vers une *GPD* passe par la détermination d'un seuil adéquat (pas très bas et pas très haut). Le choix du seuil doit être un compromis de sorte que le seuil déterminé soit suffisamment grand pour pouvoir utiliser les résultats asymptotiques, mais pas trop élevé non plus, afin de garder un nombre d'excès suffisant pour estimer convenablement les paramètres du modèle. Généralement, le seuil μ est déterminé graphiquement par le graphe de la fonction moyenne des excès (Mean Excess Plot).

Le graphe de la fonction moyenne des excès

A l'aide de la fonction *mePlot()* qu'on trouve dans le package **fExtremes** on obtient le graphe suivant :

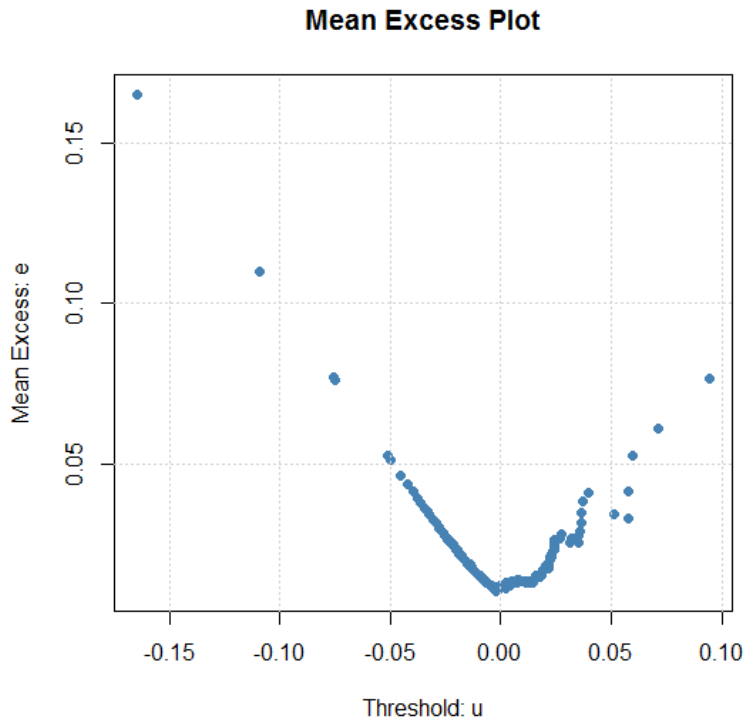


FIGURE 4.6 – La distribution moyenne des excès.

Intrprétation

D'après le graphe, le seuil $u_0 = 0,03$ tel que le nuage de points $(u, e(u))$ soit approximativement linéaire pour $u > u_0$.

1. Estimation des paramètres de la GPD par la méthode de maximum de vraisemblance

Dans cette partie, nous déterminons des estimations des paramètres de forme ξ et d'échelle σ de la GPD, en utilisant le logiciel R et les méthodes du maximum de vraisemblance.

A l'aide de la fonction `gpd()` qu'on trouve dans le package `evir`, nous obtenons le résultat suivant :

```
> y = gpd(Deltaprix, 0.03, method = "ml")
```

Les paramètres estimés sont : Pour $\mu = 0.03$ le nombre d'observation des excès $N_\mu = 16$.

$\hat{\xi}$	$\hat{\sigma}$
0.366053969	0.006791669

TABLE 4.3 – Les paramètres obtenus

On trace la courbe de la GPD avec les paramètres estimés ξ et σ :

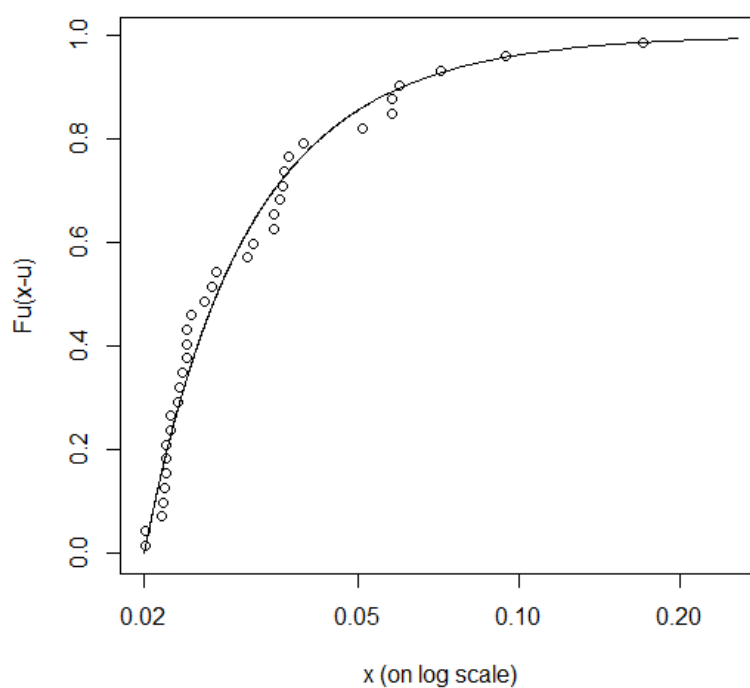


FIGURE 4.7 – L'approximation des excès par une GPD.

Interprétation

D'après la figure 4.7, on remarque que la distribution des excès converge vers une distribution de Pareto Généralisée (GPD).

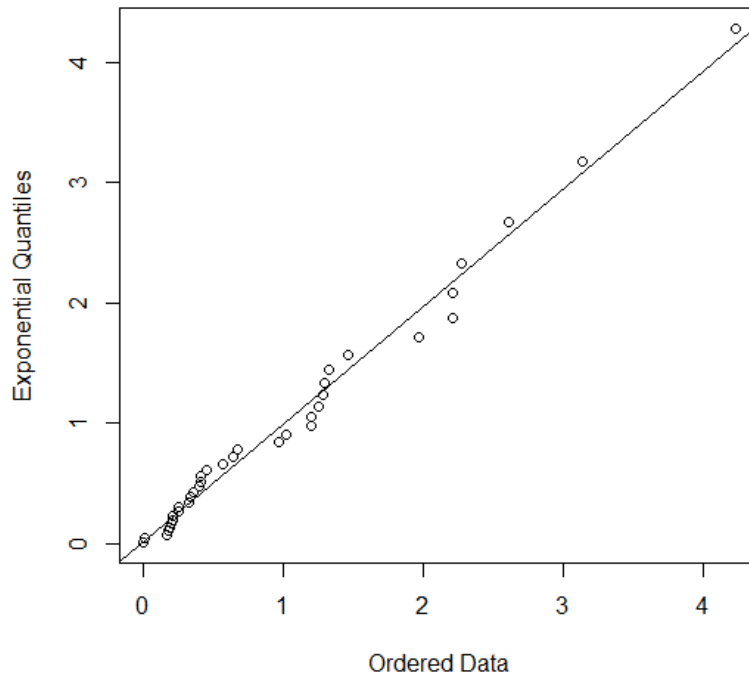


FIGURE 4.8 – QQ-plot.

Interprétation

On constate qu'il y a une linéarité des quantiles empiriques et les quantiles de la GPD estimée, ce qui nous permet de dire que l'échantillon des excès suit une GPD (0.366053969,0.006791669).

La loi GPD avec les paramètres estimés

$$G_{\xi,\sigma} = 1 - \left(1 + \frac{0.366053969}{0.006791669}x\right)^{\frac{-1}{0.366053969}} \quad \text{si } \xi \neq 0$$

2. Estimation de la VaR par la méthode du maximum de vraisemblance au seuil $u_0 = 0.03$

La VaR par la méthode des valeurs extrêmes est calculé avec la fonction `riskmeasures()` qu'on trouve dans le package `evir`. :

```
> riskmeasures(y, 0.95)
  p    quantile
0.95 0.02655226
```

Le quantile désigne la VaR.

Cela signifie que il y a 95% de chances que la perte associée à la détention de l'actif ne dépasse pas 0.02583465.

4.8 Comparaison

Dans cette partie, on s'intéresse à la comparaison des méthodes d'estimation de l'action boursière KVV qui cotée à la bourse d'amsterdam.

L'indice	Méthode	VaR(95%)
KVV	VaR Normale	1.677699
	VaR Historique	0.02223679
	VaR TVE	0.02655226

TABLE 4.4 – La VaR pour les différentes méthodes de niveau 0.95

D'après le tableau (4.4) on constate que :

- ★ La VaR TVE donne un résultat approximativement le même que la VaR historique.
- ★ La VaR normale donne un résultat éloigné des autres méthodes cela est dû à la loi des excès qui ne suit pas une loi normale.

L'avantage de l'approche VaR qu'elle produit un seul chiffre qui exprime la position de risque de marché de la banque.

Conclusion générale

Ce travail à compter quatre grandes parties. La première avait pour objectif de définir la théorie des valeurs extrêmes. Dans la deuxième partie, nous avons cité les méthodes d'estimation de quantiles et les paramètres de la loi des valeurs extrêmes.

La troisième partie, consiste à définir la méthodologie de la Valeur à Risque et les paramètres liés à son calcul. Ensuite, on a donné des grandes méthodes classiques de l'estimation de la VaR.

Enfin, dans La dernière partie nous avons effectué une étude sur une action boursière KvW qui cotée a la bourse d'amsterdam sur une période de 17 avril 2018 jusqu'à 1 avril 2020. Dans le but de comparer toutes les techniques d'estimation de la Valeur à Risque.

Il n'est pas possible d'identifier une méthode universellement applicable pour le calcul de la VaR, chaque méthode présente c'est propre limite. Nous pouvons déduire que la méthode de simulation historique est indépendante de l'hypothèse de normalité des rendements. Cette méthode d'estimation donne de meilleurs résultats lorsqu'il s'agit de grands historiques. Cependant, La méthode des variances-covariances est meilleure lorsqu'il s'agit des rendements qui suivent une loi normale.

Enfin, La méthode des valeurs extrêmes peut être considérée comme un supplément pour les méthodes d'estimation classique de la VaR. Cette dernières est applicable dans le cas des distributions ayant des queues épaisses.

Bibliographie

- [1] Amine Amar. Apport de la théorie des valeurs extrêmes à la modélisation et la gestion des risques boursiers, financiers et hydro-météorologiques. 2014.
- [2] Soufia Amrani. *Estimation de la valeur à risque en finance*. PhD thesis, Université du Québec à Trois-Rivières, 2002.
- [3] Adelchi Azzalini. A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, 68(1) :326–328, 1981.
- [4] August A Balkema and Laurens De Haan. Residual life time at great age, the annals of probability. 1974.
- [5] Adrian W Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2) :353–360, 1984.
- [6] Song Xi Chen and Cheng Yong Tang. Nonparametric inference of value-at-risk for dependent financial returns. *Journal of financial econometrics*, 3(2) :227–255, 2005.
- [7] Anthony C Davison and Richard L Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society : Series B (Methodological)*, 52(3) :393–425, 1990.
- [8] Jan G De Gooijer, Ali Gannoun, and Dawit Zerom. Mean squared error properties of the kernel-based multi stage median predictor for time series. *Statistics and probability letters*, 56(1) :51–56, 2002.
- [9] Arnold LM Dekkers, Laurens De Haan, et al. On the estimation of the extreme-value index and large quantile estimation. *The annals of statistics*, 17(4) :1795–1832, 1989.
- [10] Jean-Fran is Delmas and Benjamin Jourdain. *Modèles Aléatoires*, volume 57. Springer, 2006.
- [11] Kevin Dowd. *Measuring market risk*. John Wiley & Sons, 2007.
- [12] Jonathan El Methni. *Contributions à l'estimation de quantiles extrêmes. Applications à des données environnementales*. PhD thesis, 2013.
- [13] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events : for insurance and finance*, volume 33. Springer Science & Business Media, 2013.

-
- [14] Ronald Aylmer Fisher and Leonard Henry Caleb Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge University Press, 1928.
- [15] Raoul Fokou. *Mesure du risque de marché d'un portefeuille de type Actions (Value-At-Risk, Value-At-Risk Conditionnelle)*. PhD thesis, PhD thesis, Euro Institut d'Actuariat, 2006.
- [16] Maurice Fréchet. Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Math. Polon.*, 6 :93–116, 1927.
- [17] Ali Gannoun, Stéphane Girard, Christiane Guinot, and Jérôme Saracco. Trois méthodes non paramétriques pour l'estimation de courbes de référence-application à l'analyse de propriétés biophysiques de la peau. *Revue de Statistique appliquée*, 50(1) :65–89, 2002.
- [18] Ali Gannoun, Jérôme Saracco, and Keming Yu. Nonparametric prediction by conditional median and quantiles. *Journal of statistical Planning and inference*, 117(2) :207–223, 2003.
- [19] Boris Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.
- [20] Maurizio Guida and Maurizio Longo. Estimation of probability tails based on generalized extreme value distributions. *Reliability Engineering and System Safety*, 20(3) :219–242, 1988.
- [21] M Hervé. Aide-memoire de statistique appliquee a la biologie—construire son etude et analyser les resultats a l'aide du logiciel r. *Version 0.9-64 <https://cran.r-project.org/web/packages/RVAideMemoire>*, 2016.
- [22] Bruce M Hill. A simple general approach to inference about the tail of a distribution. *The annals of statistics*, pages 1163–1174, 1975.
- [23] Jonathan Richard Morley Hosking, James R Wallis, and Eric F Wood. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3) :251–261, 1985.
- [24] Marie Kratz and Sidney I Resnick. The qq-estimator and heavy tails. *Stochastic Models*, 12(4) :699–724, 1996.
- [25] Alexandre Lekina. *Estimation non-paramétrique des quantiles extrêmes conditionnels*. PhD thesis, 2010.
- [26] Simone Manganelli and Robert F Engle. Value at risk models in finance. 2001.
- [27] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management : concepts, techniques and tools-revised edition*. Princeton university press, 2015.
- [28] Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1) :141–142, 1964.

-
- [29] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3) :1065–1076, 1962.
- [30] Emanuel Parzen. Nonparametric statistical data modeling. *Journal of the American statistical association*, 74(365) :105–121, 1979.
- [31] Karl Pearson. V. on the methematical theory of errors of judgement, with special reference to the personal equation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 198(300-311) :235–299, 1902.
- [32] James Pickands III et al. Statistical inference using extreme order statistics. *the Annals of Statistics*, 3(1) :119–131, 1975.
- [33] P Prescott and AT Walden. Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67(3) :723–724, 1980.
- [34] Bechir Raggad. Fondements de la théorie des valeurs extrêmes, ses principales applications et son apport à la gestion des risques du marché pétrolier. *Mathématiques et sciences humaines. Mathematics and social sciences*, (186) :29–63, 2009.
- [35] Sidney I Resnick. *Extreme values, regular variation and point processes*. Springer, 2013.
- [36] Mats Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, pages 65–78, 1982.
- [37] Hanene Ben Salah. *Gestion des actifs financiers : de l'approche Classique à la modélisation non paramétrique en estimation du DownSide Risk pour la constitution d'un portefeuille efficient*. PhD thesis, 2015.
- [38] Khaled Salhi. *Risques extrêmes en finance : analyse et modélisation*. PhD thesis, 2016.
- [39] J Schultze and Josef Steinebach. On least squares estimates of an exponential tail coefficient. *Statistics and Risk Modeling*, 14(4) :353–372, 1996.
- [40] David W Scott and George R Terrell. Biased and unbiased cross-validation in density estimation. *Journal of the american Statistical association*, 82(400) :1131–1146, 1987.
- [41] Richard L Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1) :67–90, 1985.
- [42] Richard L Smith. Estimating tails of probability distributions. *The annals of Statistics*, pages 1174–1207, 1987.
- [43] Geoffrey S Watson. Smooth regression analysis. *Sankhyā : The Indian Journal of Statistics, Series A*, pages 359–372, 1964.
- [44] S Weinstein. Theory and application of some classical and generalized asymptotic distributions of extreme values. *IEEE Transactions on Information Theory*, 19(2) :148–154, 1973.
- [45] Ishay Weissman. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364) :812–815, 1978.

- [46] Shie-Shien Yang. A smooth nonparametric estimator of a quantile function. *Journal of the American Statistical Association*, 80(392) :1004–1011, 1985.
- [47] N Zougab. Etude comparative des méthodes de sélection du parametre de lissage dans l'estimation de la densité de probabilité par la méthode du noyau. *Mémoire de Magister. Université de Béjaia, Algérie*, 2007.

Annexe

Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close
17/04/2018	20,560995	17/05/2018	20,99151	15/06/2018	20,805418	16/07/2018	17,744154	14/08/2018	16,767155	12/09/2018	16,627584
18/04/2018	20,578999	18/05/2018	20,935684	18/06/2018	20,768198	17/07/2018	17,790678	15/08/2018	16,832289	13/09/2018	16,450794
19/04/2018	20,615005	21/05/2018	20,972902	19/06/2018	20,842634	18/07/2018	17,67902	16/08/2018	16,646193	14/09/2018	16,702023
20/04/2018	20,615005	22/05/2018	21,01012	20/06/2018	20,563494	19/07/2018	17,781374	17/08/2018	16,515928	17/09/2018	16,664803
23/04/2018	20,615005	23/05/2018	20,917072	21/06/2018	20,470446	20/07/2018	17,688326	20/08/2018	16,534536	18/09/2018	16,785767
24/04/2018	20,470974	24/05/2018	20,935684	22/06/2018	20,265743	23/07/2018	17,651108	21/08/2018	16,841595	19/09/2018	16,692717
25/04/2018	20,380947	25/05/2018	21,04734	25/06/2018	20,637932	24/07/2018	16,999777	22/08/2018	17,251003	20/09/2018	16,953251
26/04/2018	20,506983	28/05/2018	20,917072	26/06/2018	20,507666	25/07/2018	16,850899	23/08/2018	17,353355	21/09/2018	16,925337
27/04/2018	20,651016	29/05/2018	20,228523	27/06/2018	20,321571	26/07/2018	17,046299	24/08/2018	17,223089	24/09/2018	17,269613
30/04/2018	21,191149	30/05/2018	20,321571	28/06/2018	20,470446	27/07/2018	17,064909	27/08/2018	17,381271	25/09/2018	17,036993
02/05/2018	21,137133	31/05/2018	20,191303	29/06/2018	20,489056	30/07/2018	17,074213	28/08/2018	16,702023	26/09/2018	16,999777
03/05/2018	21,281168	01/06/2018	20,154085	02/07/2018	20,358788	31/07/2018	17,167261	29/08/2018	17,018385	27/09/2018	17,074213
04/05/2018	21,317177	04/06/2018	20,284351	03/07/2018	20,135473	01/08/2018	17,185869	30/08/2018	17,706936	28/09/2018	16,841595
07/05/2018	21,717283	05/06/2018	20,265743	04/07/2018	18,05121	02/08/2018	17,055605	31/08/2018	17,223089	01/10/2018	16,999777
08/05/2018	21,810329	06/06/2018	20,824026	05/07/2018	17,846506	03/08/2018	17,157957	03/09/2018	17,334745	02/10/2018	16,897423
09/05/2018	21,680063	07/06/2018	20,71237	06/07/2018	17,865118	06/08/2018	17,102127	04/09/2018	17,148651	03/10/2018	16,841595
10/05/2018	21,512577	08/06/2018	20,582104	09/07/2018	18,05121	07/08/2018	17,120737	05/09/2018	16,832289	04/10/2018	16,748545
11/05/2018	21,400919	11/06/2018	20,377398	10/07/2018	18,05121	08/08/2018	17,148651	06/09/2018	16,664803	05/10/2018	16,506622
14/05/2018	21,307873	12/06/2018	20,470446	11/07/2018	18,023298	09/08/2018	17,102127	07/09/2018	16,590364	08/10/2018	16,432184
15/05/2018	21,605623	13/06/2018	20,61932	12/07/2018	17,809288	10/08/2018	17,027689	10/09/2018	16,488012	09/10/2018	16,525232
16/05/2018	21,456747	14/06/2018	20,861246	13/07/2018	17,632498	13/08/2018	16,925337	11/09/2018	16,24609	10/10/2018	15,697111
Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close
11/10/2018	15,166739	09/11/2018	15,055083	10/12/2018	12,242028	11/01/2019	13,61594	11/02/2019	14,051802	12/03/2019	17,851377
12/10/2018	15,0923	12/11/2018	14,93412	11/12/2018	12,971622	14/01/2019	13,511712	12/02/2019	14,392911	13/03/2019	17,671349
15/10/2018	14,859682	13/11/2018	14,72011	12/12/2018	13,265355	15/01/2019	13,426435	13/02/2019	14,364485	14/03/2019	17,851377
16/10/2018	15,14813	14/11/2018	14,943425	13/12/2018	13,748593	16/01/2019	13,568563	14/02/2019	14,506614	15/03/2019	17,879807
17/10/2018	15,203958	15/11/2018	13,864074	14/12/2018	13,454861	17/01/2019	13,59699	15/02/2019	14,828772	18/03/2019	18,116688
18/10/2018	15,0923	16/11/2018	14,357226	17/12/2018	13,644365	18/01/2019	13,843346	18/02/2019	15,084604	19/03/2019	18,078785
19/10/2018	14,906207	19/11/2018	14,468884	18/12/2018	13,720167	21/01/2019	13,81492	19/02/2019	15,188832	20/03/2019	18,021935
22/10/2018	14,580541	20/11/2018	14,478188	19/12/2018	13,39801	22/01/2019	13,919147	20/02/2019	16,695398	21/03/2019	18,088261
23/10/2018	14,310703	21/11/2018	14,572941	20/12/2018	13,009523	23/01/2019	13,691742	21/02/2019	16,335337	22/03/2019	17,443943
24/10/2018	14,571235	22/11/2018	14,307634	21/12/2018	13,0569	24/01/2019	13,720167	22/02/2019	16,62907	25/03/2019	17,406042
25/10/2018	14,664283	23/11/2018	14,203405	24/12/2018	12,801068	25/01/2019	13,824395	25/02/2019	16,477467	26/03/2019	17,737675
26/10/2018	14,282789	26/11/2018	14,53504	27/12/2018	12,592613	28/01/2019	13,833872	26/02/2019	16,439566	27/03/2019	17,519743
29/10/2018	14,599149	27/11/2018	14,497139	28/12/2018	12,990573	29/01/2019	13,947574	27/02/2019	16,676447	28/03/2019	17,415516
30/10/2018	14,72011	28/11/2018	14,137079	31/12/2018	13,104277	30/01/2019	14,070752	28/02/2019	17,292339	29/03/2019	17,415516
31/10/2018	14,608454	29/11/2018	13,928623	02/01/2019	13,085326	31/01/2019	14,19393	01/03/2019	17,339714	01/04/2019	17,434467
01/11/2018	14,841073	30/11/2018	13,919147	03/01/2019	12,971622	01/02/2019	14,279207	04/03/2019	18,249338	02/04/2019	17,605022
02/11/2018	14,822464	03/12/2018	13,805445	04/01/2019	13,265355	04/02/2019	14,269732	05/03/2019	18,135635	03/04/2019	17,813477
05/11/2018	14,887597	04/12/2018	13,625416	07/01/2019	13,559089	05/02/2019	14,497139	06/03/2019	18,135635	04/04/2019	17,642923
06/11/2018	15,241177	05/12/2018	13,388534	08/01/2019	13,729643	06/02/2019	14,686645	07/03/2019	17,984032	05/04/2019	17,472368
07/11/2018	15,455187	06/12/2018	13,085326	09/01/2019	13,909674	07/02/2019	14,307634	08/03/2019	17,567122	08/04/2019	17,491318
08/11/2018	15,14813	07/12/2018	13,208504	10/01/2019	13,79597	08/02/2019	14,042326	11/03/2019	17,74715	09/04/2019	17,244963

Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close
10/04/2019	17,320765	14/05/2019	17,161285	12/06/2019	17,951218	11/07/2019	17,891973	09/08/2019	17,556252	09/09/2019	15,008718
11/04/2019	17,699774	15/05/2019	17,141539	13/06/2019	17,990713	12/07/2019	18,089457	12/08/2019	17,635246	10/09/2019	15,087712
12/04/2019	17,586071	16/05/2019	17,220531	14/06/2019	17,714239	15/07/2019	18,010462	13/08/2019	17,793232	11/09/2019	15,107461
15/04/2019	18,021935	17/05/2019	17,062544	17/06/2019	17,477259	16/07/2019	17,990713	14/08/2019	17,299524	12/09/2019	15,028468
16/04/2019	18,287239	20/05/2019	16,707075	18/06/2019	17,951218	17/07/2019	17,931469	15/08/2019	17,200783	13/09/2019	15,265448
17/04/2019	18,495695	21/05/2019	16,687325	19/06/2019	17,911172	18/07/2019	17,45751	16/08/2019	17,575998	16/09/2019	15,601169
18/04/2019	18,021935	22/05/2019	16,667576	20/06/2019	17,654993	19/07/2019	17,418013	19/08/2019	17,832727	17/09/2019	15,46293
23/04/2019	18,059834	23/05/2019	16,272612	21/06/2019	17,635246	22/07/2019	17,674742	20/08/2019	17,852476	18/09/2019	15,423433
24/04/2019	18,346184	24/05/2019	16,391102	24/06/2019	17,260027	23/07/2019	17,497005	21/08/2019	18,148701	19/09/2019	15,541924
25/04/2019	18,425177	27/05/2019	16,410851	25/06/2019	17,220531	24/07/2019	17,595749	22/08/2019	17,832727	20/09/2019	15,719659
26/04/2019	18,385681	28/05/2019	16,252863	26/06/2019	17,358768	25/07/2019	17,654993	23/08/2019	17,911172	23/09/2019	15,640665
29/04/2019	18,721401	29/05/2019	15,956638	27/06/2019	17,378517	26/07/2019	17,891973	26/08/2019	18,207947	24/09/2019	15,58142
30/04/2019	18,820143	30/05/2019	16,015882	28/06/2019	17,437761	29/07/2019	17,69449	27/08/2019	18,563416	25/09/2019	15,46293
02/05/2019	18,385681	31/05/2019	16,05538	01/07/2019	17,575998	30/07/2019	17,556252	28/08/2019	18,484423	26/09/2019	15,69991
03/05/2019	18,28694	03/06/2019	15,8184	02/07/2019	17,654993	31/07/2019	17,793232	29/08/2019	15,680161	27/09/2019	15,8184
06/05/2019	18,030212	04/06/2019	16,134373	03/07/2019	17,714239	01/08/2019	18,148701	30/08/2019	15,403686	30/09/2019	16,035631
07/05/2019	17,773483	05/06/2019	16,312109	04/07/2019	17,635246	02/08/2019	17,891973	02/09/2019	15,048216	01/10/2019	16,331858
08/05/2019	17,852476	06/06/2019	16,707075	05/07/2019	18,207947	05/08/2019	17,319273	03/09/2019	14,890229	02/10/2019	15,838149
09/05/2019	16,983551	07/06/2019	17,33902	08/07/2019	17,773483	06/08/2019	17,575998	04/09/2019	14,969222	03/10/2019	15,69991
10/05/2019	17,615498	10/06/2019	17,556252	09/07/2019	17,497005	07/08/2019	17,378517	05/09/2019	15,008718	04/10/2019	15,69991
13/05/2019	16,884809	11/06/2019	17,911172	10/07/2019	17,753735	08/08/2019	17,654993	06/09/2019	14,988971	07/10/2019	16,865061
Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close	Date	Adj Close
08/10/2019	16,312109	06/11/2019	20,785101	05/12/2019	22	08/01/2020	21,9	06/02/2020	21,950001	06/03/2020	21,950001
09/10/2019	16,549088	07/11/2019	20,735731	06/12/2019	21,950001	09/01/2020	21,950001	07/02/2020	21,950001	09/03/2020	21,950001
10/10/2019	16,667576	08/11/2019	20,834473	09/12/2019	22	10/01/2020	21,9	10/02/2020	21,9	10/03/2020	21,950001
11/10/2019	17,260027	11/11/2019	20,686359	10/12/2019	22	13/01/2020	21,950001	11/02/2020	21,950001	11/03/2020	21,950001
14/10/2019	17,141539	12/11/2019	21,92063	11/12/2019	21,950001	14/01/2020	21,9	12/02/2020	22	12/03/2020	21,950001
15/10/2019	17,358768	13/11/2019	21,969999	12/12/2019	21,9	15/01/2020	21,950001	13/02/2020	21,950001	13/03/2020	21,950001
16/10/2019	17,299524	14/11/2019	21,92063	13/12/2019	21,950001	16/01/2020	21,9	14/02/2020	21,9	16/03/2020	21,950001
17/10/2019	17,477259	15/11/2019	21,92063	16/12/2019	22	17/01/2020	21,950001	17/02/2020	21,950001	17/03/2020	21,950001
18/10/2019	17,200783	18/11/2019	21,92063	17/12/2019	21,9	20/01/2020	21,9	18/02/2020	21,9	18/03/2020	21,950001
21/10/2019	17,200783	19/11/2019	21,969999	18/12/2019	22	21/01/2020	21,950001	19/02/2020	21,9	19/03/2020	21,950001
22/10/2019	17,358768	20/11/2019	21,950001	19/12/2019	21,9	22/01/2020	21,950001	20/02/2020	21,950001	20/03/2020	21,950001
23/10/2019	17,418013	21/11/2019	21,9	20/12/2019	21,950001	23/01/2020	21,9	21/02/2020	21,9	23/03/2020	21,9
24/10/2019	17,220531	22/11/2019	21,9	23/12/2019	21,9	24/01/2020	21,950001	24/02/2020	21,950001	24/03/2020	21,950001
25/10/2019	17,220531	25/11/2019	21,9	24/12/2019	21,950001	27/01/2020	21,950001	25/02/2020	21,9	25/03/2020	21,950001
28/10/2019	17,477259	26/11/2019	22	27/12/2019	21,9	28/01/2020	21,9	26/02/2020	21,9	26/03/2020	21,950001
29/10/2019	20,735731	27/11/2019	21,9	30/12/2019	21,950001	29/01/2020	21,9	27/02/2020	21,950001	27/03/2020	22,1
30/10/2019	20,834473	28/11/2019	21,9	31/12/2019	21,950001	30/01/2020	21,9	28/02/2020	21,9	30/03/2020	22,15
31/10/2019	20,538246	29/11/2019	21,950001	02/01/2020	21,950001	31/01/2020	21,9	02/03/2020	22	31/03/2020	21,700001
01/11/2019	20,735731	02/12/2019	21,950001	03/01/2020	21,9	03/02/2020	21,950001	03/03/2020	21,950001		
04/11/2019	20,735731	03/12/2019	21,9	06/01/2020	21,950001	04/02/2020	21,950001	04/03/2020	21,950001		
05/11/2019	20,785101	04/12/2019	22	07/01/2020	21,950001	05/02/2020	21,950001	05/03/2020	21,950001		

FIGURE 4.9 – Les prix de clôture de l'action KvW en Euros.

Résumé

Dans ce travail, nous avons présenté dans un premier temps, quelques rappels et définitions sur la statistique d'ordre et le quantile en mentionnant les principaux résultats de la théorie des valeurs extrêmes. En deuxième lieu, nous avons fait appel à des méthodes pour estimer le quantile et l'indice des valeurs extrêmes. La troisième partie, consiste à définir la méthodologie de la Valeur à Risque, puis on a cité des grandes méthodes classiques d'estimation de ce dernier. Une application en finance est donnée pour estimer le quantile extrême et la valeur à risque suivant les différentes méthodes.

Mots-clés : Valeurs extrêmes, Statistique d'ordre, Quantile extrême, Valeur à risque.

Abstract

In this work, we first presented some reminders and definitions on order statistics and the quantile, mentioning the main results of the theory of extreme values. In a second, we used methods to estimate the quantile and the index of extreme values. The third part consists in defining the methodology of Value at Risk, then we have cited the major classical methods of estimating the latter. An application in finance is given to estimate the extreme quantile and the value at risk according to the different methods.

Keywords : Extreme values, Order statistics, Extreme quantile, Value at risk.