

THÈSE DE DOCTORAT

Présentée par

Amine AMAR

Discipline : Mathématiques Appliquées

Spécialité : Statistiques

Apport de la théorie des valeurs extrêmes à la modélisation et la gestion des risques boursiers, financiers et hydro-météorologiques

Soutenue le 01/03/2014 devant le jury :

Président :

Pr. Khalid EL HIMDI

Professeur à la Faculté des Sciences,
Université Mohammed V-Agdal, Rabat.

Examineurs :

Pr. Zine El Abidine GUENNOUN

Professeur à la Faculté des Sciences,
Université Mohammed V- Agdal, Rabat.

Pr. Mohammedine BELBACHIR

Professeur à la Faculté des Sciences,
Université Mohammed V- Agdal, Rabat.

Pr. Abdelaziz CHAOUBI

Professeur à l'Institut National de Statistique et
d'Economie Appliquée (INSEA), Rabat.

Pr. Elhadj EZZAHID

Professeur Habilité à la Faculté des Sciences
Juridiques, Economiques et Sociales, Université
Mohammed V-Agdal, Rabat.

Pr. Abdelhak ZOGLAT

Professeur à la Faculté des Sciences,
Université Mohammed V- Agdal, Rabat.

Avant propos

Ce travail a été réalisé au Laboratoire de Mathématiques Appliquées de la Faculté des Sciences de l'Université Mohammed V-Agdal, sous la direction de Monsieur Abdelhak ZOGLAT, professeur d'enseignement supérieur et responsable dudit laboratoire.

La thèse a été effectuée sous le co-encadrement de Monsieur Salah-eddine El ADLOUNI, Professeur à l'Université de Moncton au Canada.

Au terme de ce travail, je tiens tout d'abord à exprimer ma gratitude, mon respect et mes remerciements à mon directeur de thèse, Pr. Abdelhak ZOGLAT. Je ne serais pas arrivé jusque là sans sa disponibilité et son investissement constant. Je remercie vivement mon co-encadrant Pr. Salah-eddine El ADLOUNI qui m'a donné l'opportunité d'étudier le comportement des valeurs extrêmes, pour la qualité de ses conseils précieux et perspicaces. Je remercie tous les membres du laboratoire qui m'ont permis de m'épanouir aussi bien dans mes recherches que dans mes enseignements.

J'adresse mes remerciements plus particulièrement au Pr. Khalid EL HIMDI, professeur à la Faculté des Sciences, Université Mohammed V- Agdal, Rabat, président du jury et examinateur de ma thèse. Merci de m'avoir honoré en acceptant de présider le jury de la thèse.

Je tiens à remercier les membres de mon comité de thèse, pour l'intérêt qu'ils ont porté à ce travail et pour les remarques constructives qu'ils ont fait sur mon projet de thèse.

Je remercie Pr. Mohammedine BELBACHIR, professeur à la Faculté des Sciences, Université Mohammed V- Agdal, Rabat, et examinateur de ma thèse, pour ses excellentes corrections, qui m'ont permis d'améliorer la qualité de mon travail de recherche.

Je remercie Pr. Zine El Abidine GUENNOUN, professeur à la Faculté des Sciences, Université Mohammed V- Agdal, Rabat, rapporteur et examinateur de ma thèse, pour ses conseils et ses directives perspicaces.

Je remercie Pr. Abdelaziz CHAOUBI, professeur à l'Institut National de Sta-

tistique et d'Economie Appliquée (INSEA), Rabat, rapporteur et examinateur de ma thèse, pour ses remarques et ses encouragements.

Je remercie Pr. Elhadj EZZAHID, professeur habilité à la Faculté des Sciences Juridiques, Economiques et Sociales, Université Mohammed V-Agdal, Rabat, et examinateur de ma thèse, pour ses suggestions considérables.

Un merci tout particulier au personnel administratif de la Faculté des Sciences de Rabat, Agdal. En particulier, le personnel du Département de Mathématiques et du Centre d'Etudes Doctorales.

Dédicaces

Je tiens à manifester toute mon affection et ma gratitude à ma mère, ma grande mère, ma soeur, mes tantes et mon oncle pour le soutien sans failles.

Un remerciement tout particulier à ma Chère épouse Fatime Zahrae, merci pour tes précieuses prières.

Je remercie ensuite tous les membres de ma famille. Merci d'avoir supporté ma mauvaise humeur le long de toutes ces années de thèse et j'espère, pouvoir rattraper mon absence fréquente. Merci à mes amis éternels Meryem, Chaouki, Adil et Bessam, de m'avoir épaulé pendant les moments les plus ternes.

Je profite de l'occasion aussi, pour remercier les charmantes basma, Leila et Malak et le Charmant Karim. Merci d'avoir apporté le bonheur à toute la famille.

Je remercie également le Grand Cyrille et la collègue Fadoua, avec qui j'ai partagé les bons et mauvais moments de ces années de thèse. Je tiens à leur dire que ce travail est le fruit de leurs accompagnements et encouragements.

Je ne manque pas de remercier, pour les tasses de thé offertes, le personnel du café où j'ai passé les plus beaux moments d'inspiration.

Enfin, je souhaite dédier ce travail à la mémoire de mon grand père, mon père et ma tante Mina. Je tiens à remercier tous ceux qui n'ont pas douté en aucun moment, que je puisse franchir la ligne d'arrivée.

Table des matières

1	Approche des maxima par blocs : lois GEV	19
1	Valeurs extrêmes et lois α -stables	19
2	Théorème fondamental des valeurs extrêmes	22
3	Estimation des modèles GEV	27
3.1	Estimation pour la loi de Gumbel	28
3.2	Estimation pour les lois de Fréchet et de Weibull	30
4	Non stationnarité et dépendance	31
4.1	Les extrêmes dans le cas des suites d'observations dé- pendantes et stationnaires :	32
4.2	GEV non stationnaires :	34
2	Approche des dépassements de seuil : loi GPD	37
1	Modélisation des excès	37
2	Méthodes pour la détection du seuil	40
2.1	Méthodes graphiques	40
2.2	Méthodes numériques	41
3	Estimation du modèle GPD	42
3.1	Méthode des moments	42
3.2	Méthode du maximum de vraisemblance	42
3.3	Méthode des moments de probabilité pondérés	42
4	Méthodes non paramétriques pour l'estimation de ξ	43
4.1	L'estimateur de <i>Pickands</i>	43
4.2	L'estimateur de <i>Hill</i>	43
4.3	L'estimateur des moments de <i>Dekkers – Einmahl – De Haan</i>	44
5	GPD dans le cas des séries non stationnaires	44
5.1	GPD non stationnaire : paramétrisation de la non sta- tionnarité	45
5.2	GPD non stationnaire : définition d'un modèle de mélange	45

3	Adéquation des modèles des valeurs extrêmes	49
1	Introduction	49
2	Probabilité et Quantile Plot	49
3	Tests usuels d'adéquation	50
4	Test de rapport de vraisemblance	50
5	Exponential Tail "ET" test	52
4	Copules extrêmes	55
1	Présentation mathématique des copules	58
1.1	Théorème de Sklar et propriétés des copules	58
1.2	Exemples de copules	59
1.3	Formes de dépendance	62
2	Les copules extrêmes	65
3	Les copules archimax	66
4	Méthodes d'estimation	67
4.1	Les copules empiriques	67
4.2	Estimation par la méthode du maximum de vraisemblance	67
5	Modélisation des extrêmes pour la gestion des risques hydro- météorologiques	69
1	Gestion des risques hydro-météorologiques	69
2	Application du POT aux données hydrologiques	71
2.1	Introduction	71
2.2	Données	71
2.3	Méthodologies	74
2.4	Résultats	80
6	Gestion des mouvements extrêmes : Crises financières, crashes boursiers et chocs pétroliers	85
1	Mouvements extrêmes liés aux marchés boursiers	85
2	Chocs pétroliers et impacts économiques	87
3	Gestion des mouvements extrêmes	88
4	Application of POT approach to CAC40 return index	91
4.1	Introduction	91
4.2	Données	92
4.3	Méthodologies	94
4.4	Résultats	94

Résumé

Le risque est un phénomène inhérent à plusieurs domaines qui relèvent de la vie des décideurs et des investisseurs. A ce titre, les chercheurs portent un intérêt très particulier à l'évaluation, la gestion et à la prévision des différents types de risques, pouvant entraver le bon déroulement d'une activité ou menacer un projet ou une infrastructure. C'est dans ce contexte que la présente thèse s'inscrit. Nous appliquons les techniques de la théorie des valeurs extrêmes (TVE) à la gestion des risques hydrologiques et boursiers.

Sur le plan organisationnel, la thèse se décompose en deux grandes parties distinctes. Dans un premier temps, nous nous intéressons aux fondements de la théorie des valeurs extrêmes, en mettant l'accent sur l'approche "Peaks Over Threshold (POT)". Nous nous focalisons sur l'étape cruciale de cette approche qui réside dans le choix d'un seuil optimal. Nous proposons la comparaison de plusieurs approches objectives basées sur des méthodes statistiques développées. Ce premier bloc qui constitue la partie théorique de notre thèse, présente également différentes extensions de la théorie classique, à savoir les copules extrêmes, les GEV non-stationnaires et le traitement des extrêmes dans le cas d'observations dépendantes. La deuxième partie, à vocation pratique, est consacrée aux applications et au traitement des données réelles. Ces applications illustrent l'utilité de la TVE pour une meilleure modélisation et gestion des risques. Enfin, nous avons proposé en conclusion, des perspectives et des pistes de recherche dont les fruits sont encore en cours d'exploration et d'achèvement.

Mots Clés : Automated Threshold Selection Method, Generalized Extreme Value Distributions, Generalized Pareto Distribution, Likelihood Ratio Test, Mean Square Error Method, Multiple Threshold Method, Peaks Over Threshold, Tests d'ajustement, Risques, Théorie des Valeurs Extrêmes.

Abstract

The present thesis is placed in the context of financial and hydrological risk management. In general, risks are described as inherent phenomena of current life of decision makers and investors. Because of this reason, risks bear a very special interest of researchers.

According the organizational level, the thesis is divided into two parts. Firstly, we are interested in foundations of extreme values theory (EVT) and especially, "Peaks Over Threshold (POT)" approach. The critical step of POT lies on the choice of an optimal threshold. So, we propose in this thesis a comparison of several approaches based on objective measures and statistical methods. This first block constitutes the theoretical part of the thesis. Additionally, theoretical part contains different extensions of classical EVT, namely extreme copulas, non stationary GEV and the treatment of extremes in the case of dependent observations.

In a second step, another party is devoted to applications and processing of real data.

Finally, in conclusion, we proposed perspectives and tracks of research whose fruits are still under exploration and completion.

Keywords : Adjustment tests, Automated Threshold Selection Method, Extreme Value Theory, Generalized Extreme Value Distributions, Generalized Pareto Distribution, Likelihood Ratio Test, Mean Square Error Method, Multiple Threshold Method, Peaks Over Threshold, Risk.

Introduction Générale

La multiplication et les impacts socio-économiques importants des événements extrêmes (sécheresses intenses, pluies diluviennes, ouragans, tremblements de terre, inondations, crises financières, crashes boursiers, chocs pétroliers, etc.), justifient l'intérêt particulier qui s'est développé ces dernières années auprès des chercheurs et décideurs, afin de comprendre le comportement des risques extrêmes.

Les événements extrêmes sont généralement des événements rares, dont la probabilité d'apparition est très faible, mais dont les impacts sont considérables et les valeurs enregistrées sont beaucoup plus fortes ou plus faibles que celles observées habituellement. D'un point de vue statistique, l'étude des valeurs extrêmes se ramène à l'analyse des queues des lois de probabilité, en cherchant à estimer les quantiles extrêmes dépassés avec une probabilité très faible. L'apparition d'un événement extrême peut correspondre à un simple manque de chance (probabilité faible de se produire) comme elle peut correspondre à un changement de régime (changement structurel), et donc à l'émergence d'un nouveau phénomène qui va durer dans le temps. Ainsi, il est nécessaire de pouvoir bien discerner les deux types pour déceler le comportement des événements extrêmes isolés, et qui sont dus au pur hasard, afin d'étudier en particulier leur loi de probabilité et renseigner sur leur récurrence.

La protection d'un système (économique, financier, industriel, biodiversité, etc.) vis-à-vis de ces événements qui peuvent correspondre à des violations, revêt un intérêt particulier, car la prévention des risques conséquents peut efficacement contribuer à éviter plusieurs catastrophes.

Cependant, la compréhension des événements extrêmes est souvent confrontée à un problème de manque d'information, notamment lors d'une analyse statistique. Pire encore, les informations concernant les événements les plus fréquents occultent souvent celles des événements extrêmes, constituant ainsi des parasites qui les font noyer dans la masse des incidents habituels.

Par ailleurs, les méthodes statistiques classiques utilisent toutes les données observées sans discerner celles qui se rapportent aux événements extrêmes, fournissant ainsi des modèles appropriés pour décrire le comportement central, et n'apportant guère d'informations spécifiques aux valeurs extrêmes. Les

outils probabilistes traditionnels développés dans un cadre gaussien sont inadaptés à l'appréhension d'un univers, où les valeurs ne sont pas du même ordre de grandeur et sont loin d'être concentrées autour de leur moyenne, avec des variations de faible amplitude (Raggad (2009), [114]).

L'approche traditionnelle basée sur l'homme moyen de Quételet, n'est même pas capable de décrire dans une économie donnée, la loi de probabilité des revenus pour lesquels Pareto (fin 19^{ème} siècle) a pu constater qu'ils se répartissent selon une loi décroissante d'allure, entre une petite minorité de riches, une classe moyenne minoritaire, et une large majorité de pauvres.

Face à ces difficultés d'ordre conceptuel, les théoriciens se sont intéressés au comportement asymptotique de la loi du maximum. Cet intérêt se révèle indispensable, car il permet une analyse plus complète des lois de probabilité. L'analyse concerne en particulier, le comportement des valeurs extrêmes, situées en queues des lois.

Cette idée était à la base de toute une branche de la statistique qui est apparue entre 1920 et 1940, grâce à Fréchet, Fisher et Tippett, Gumbel et Gnedenko. Cette branche s'est bien développée à partir des années 50, même si le théorème principal a été énoncé par Fisher et Tippett dès 1928. Ladite branche s'appelle la Théorie des Valeurs Extrêmes (TVE). Grâce à son cadre mathématique rigoureux et capable de réaliser des extrapolations se rapportant au comportement stochastique des extrêmes, la TVE permet d'évaluer les événements rares et les pertes associées à leur apparition, en cherchant à ajuster un modèle approprié aux observations extrêmes. Ainsi, sous certaines conditions de régularité, la TVE énonce que la loi du maximum d'un ensemble de variables aléatoires, ne peut appartenir qu'à une seule famille de lois : Weibull (à support borné), Gumbel et Fréchet (à support non borné).

Sur le plan théorique, la TVE contient deux volets. Le premier concerne les lois des valeurs extrêmes (GEV : Generalized Extreme Values) et qui s'intéresse à la loi du maximum. Cette première approche est basée sur le théorème de Fisher et Tippett. Le second volet concerne les lois des excès (POT : Peaks Over threshold). Il consiste à utiliser les observations qui dépassent un certain seuil déterministe. Les différences entre ces observations et le seuil sont appelées excès. Cette deuxième approche est fondée sur la convergence vers une loi de Pareto Généralisée (GPD) assurée par le théorème de Pickands-Balkema-de Haan.

La première approche (blocs des maxima) a été très souvent critiquée parce qu'elle utilise un minimum d'information qui se rapporte à la valeur maximale

de chaque période d'observations (année, semestre ou trimestre par exemple). Cette manière de faire, pose des problèmes d'estimation surtout pour des séries courtes. Pour cette raison, la seconde approche (la méthode des excès) a été souvent préconisée, malgré le fait qu'en pratique, le choix du seuil constitue une vraie difficulté. En fait, le seuil doit être assez grand pour que l'approximation par une loi GPD soit valide, mais pas trop élevé pour garder un nombre suffisant de dépassements pour une meilleure estimation.

Par ailleurs, la littérature concernant les valeurs extrêmes est très riche. Nous pouvons citer par exemple les ouvrages de Leabetter et al. (1983, [83]) et de Resnick (1987, [117]) qui fournissent un traitement mathématique approfondi, le livre de Gumbel (1958, [58]) qui constitue une référence classique, l'ouvrage de Galambos (1985, [49]) qui propose un historique des principaux résultats, le livre de Coles (2001, [23]) et l'ouvrage de Reiss et Thomas (2001, [115]) qui combinent applications et théorie en proposant un certain nombre d'exemples pratiques.

D'un point de vue pragmatique, la TVE a contribué à la compréhension de plusieurs phénomènes. En démographie par exemple, tout un débat a été initié par Gumbel et auquel Fréchet a pris une part active, sur la notion de durée extrême de la vie humaine et sur sa mesure (Thatcher (1999, [136]), Horiuchi et Wilmoth (1998, [62])). Ceci a conduit les démographes à rechercher des modèles compatibles avec la réalité, et à répondre à la question soulevée sur l'existence d'un âge limite de longévité de l'être humain. Une telle étude est souvent utile pour les organismes de sécurité sociale afin de déterminer les coûts de traitement ou les dépenses destinées aux personnes âgées (Raggad (2009), [114]).

En hydrologie, domaine dans lequel la prévision est particulièrement importante, la TVE (El Adlouni et al. (2010, [40]), Garavaglia (2011, [52]), Guillou et al. (2006, [56])) permet de proposer un outil de description et de prédiction, dégagé des hypothèses trop arbitraires. L'efficacité de la TVE en gestion des risques hydrologiques s'est clairement confirmée, surtout avec l'intensification des retombées des changements climatiques.

En finance, où les booms et les crashes sont les phénomènes extrêmes les plus spectaculaires, la TVE a constitué une innovation par rapport aux modèles classiques (Longin (1998 et 2000 [90], [88]), Bei Xu (2008, [152])). La nécessité de se doter d'une telle théorie en finance a été reconnue depuis plusieurs années. En 1995, lors d'une conférence conjointe des chercheurs des banques centrales, le président de la Réserve Fédérale Américaine, déclarait : " L'existence d'un

prêteur de dernier recours, a pour conséquence naturelle que la charge du risque se voit d'une certaine façon, répartie lorsque survient un événement extrême. Ainsi, les banques centrales sont amenées à jouer un rôle qui correspond essentiellement à celui d'assurer une couverture contre les sinistres. (...). Selon les gestionnaires des risques, une utilisation abusive de la loi normale peut entraîner une sous-estimation des risques qui doit en contrepartie permettre une simplification significative. Selon les banques centrales, les conséquences sont encore plus importantes, car c'est souvent la queue gauche de la loi ajustée, que nous devons considérer lorsque nous définissons la position à adopter en tant que prêteur de dernier recours. Il est donc crucial d'améliorer la description de la loi des valeurs extrêmes." Dans l'hebdomadaire américain, *Business Week* (septembre 1998), il a été bien noté que : " Les hausses et les baisses extrêmes et synchronisées des marchés financiers sont rares, mais elles existent. Elles échappent aux modèles parce qu'ils ne prennent pas suffisamment en compte la probabilité d'un enchaînement de catastrophes ". Ainsi, les applications de la TVE en finance ont été développées récemment, afin que les modèles habituellement utilisés puissent mieux être configurés, pour intégrer les évolutions extrêmes.

Les champs d'application de la TVE se sont bien élargis, car son introduction est devenue une exigence dans de nombreux domaines : climatologie (Christidis et al. (2011, [21]), Zwiers et al. (2011, [154])), océanographie et écologie (Raillard (2011, [113]), Prevosto (2003, [112]), Thompson et al. (2001, [137])), environnement (Atyeo et al.(2011,[1])), sport (Einmahl et al.(2011, [43])), management (Dahan et Mendelson (2001,[27])), agro-alimentaire (Kawas et Moreira (2001,[75])), industrie (Klajnmic (2003,[79]), Tryon et Cruse (2000, [144])) et médecine (Roberts (2000, [121])).

Quel que soit le domaine d'application, les usagers de la TVE passent toujours par une phase préliminaire. Elle consiste à vérifier les hypothèses de base liées à l'approche utilisée. En fait, tous les modèles statistiques sont intimement liés à leurs hypothèses d'applicabilité, et souvent, certains modèles sont sélectionnés sur la base de la simplicité vis-à-vis de leur mise en application. D'autres, sont rejetés sous prétexte de complexité. En particulier, les praticiens de la TVE sont confrontés aux hypothèses d'indépendance, de stationnarité et d'homogénéité. Ces hypothèses sont dans la pratique difficiles à vérifier. D'une part, cette difficulté incite les théoriciens à étendre les résultats classiques de la TVE aux cas de non-stationnarité ou de saisonnalité. D'autre part, la non validité des hypothèses obligent les praticiens de la TVE, à procéder à quelques trans-

formations (par exemple, transformer les prix financiers en rendements).

Cela dit, les champs de recherche dans la TVE sont prometteurs. Certains champs sont encore en développement, alors que d'autres ne sont pas encore exploités. La présente thèse a été conçue, afin de contribuer à résoudre quelques problèmes en hydrologie et en finance, en adaptant les résultats de la théorie des valeurs extrêmes, notamment l'approche des excès (POT), à quelques cas pratiques. Toute la difficulté de cette approche réside dans le choix d'un meilleur seuil assurant la convergence vers une loi GPD. De ce fait, plusieurs méthodes ont été proposées dans les ouvrages et les articles traitant la TVE. Les méthodes évoquées peuvent être regroupées en deux classes. La première est constituée des méthodes graphiques. Ces méthodes simples à mettre en œuvre, souffrent malheureusement d'un problème de subjectivité. Dans le meilleur des cas, elles peuvent être utilisées pour identifier un intervalle de seuils candidats. La deuxième classe quant à elle, concerne les méthodes analytiques qui se basent sur des procédures objectives.

Le traitement des données analysées dans la présente thèse, a mis en exergue la non validité des hypothèses requises pour l'application de la TVE. Ainsi, il a été d'une grande utilité de procéder à quelques transformations des données de base afin d'appliquer les résultats théoriques.

La partie empirique de cette thèse, consiste à vérifier si le modèle proposé par l'approche POT peut décrire les séries étudiées. Nous avons pu identifier des modèles pouvant servir d'aide à la décision en finance et en hydrologie. Pour atteindre cet objectif, nous avons analysé un recueil de méthodes, dont certains aspects ont été améliorés, en intégrant des tests statistiques dans l'étape de la sélection des seuils candidats, et en jugeant les performances de chaque méthode sur la base de plusieurs outils de diagnostic et d'adéquation. Nous allons par ailleurs, évoquer des questions concernant les considérations théoriques et pratiques qui justifient le choix des séries hydrologiques et financières, leurs particularités et caractéristiques, ainsi que les modèles habituellement utilisés pour leur gestion.

Dans sa globalité, le présent travail est destiné à mettre en application la méthode POT et à évaluer ses performances. Par conséquent, le premier objectif assigné à la présente étude est d'analyser les différentes techniques de la sélection du seuil optimal et puis, d'identifier les modèles qui s'ajustent mieux aux données utilisées. Le deuxième objectif est de caractériser les lois des données hydrologiques et financières pour des considérations de gestion et de prise de décision. Ainsi, notre thèse sera structurée en deux grandes parties qui re-

groupent six chapitres.

Une première partie constituée des quatre premiers chapitres, a été dédiée à la présentation de la théorie des valeurs extrêmes et aux portées et résultats généraux des copules extrêmes. Nous fournissons un exposé de synthèse sur ce sujet, en y présentant des définitions et des propriétés fondamentales. Une attention particulière est accordée à l'approche POT. Nous évoquons aussi quelques prolongements et extensions de la TVE, à savoir, les extrêmes des suites de variables aléatoires dépendantes et la GEV non-stationnaire. Les quatre premiers chapitres constituent donc, la partie théorique qui sert à introduire divers concepts nécessaires pour mettre en contexte les questions de la présente thèse.

Les chapitres 5 et 6 ont fait l'objet de deux publications dans le *Journal of Hydrologic Engineering*, 2014 (Managing hydrological risks with extreme modeling : application of peaks over threshold model to the loukkos watershed) et l'*International Journal of Statistics and Economics*, 2012 (Statistical methods to expect extreme values : Application of POT, approach to CAC40 return index). Les principaux résultats des deux articles ont été disséminés, après leur parution dans les deux journaux mentionnés, à l'occasion de deux congrès : *Second Workshop on the Application of Stochastic Models in Finance and in Risks Management* (Rabat, 2012) et *5th International Perspective on Water Resources and the Environment Conference* (Marrakech, 2012). Les deux chapitres 5 et 6 constituent la partie empirique de la thèse. Nous avons réalisé les deux chapitres de cette partie, en deux étapes. Il s'agit dans un premier temps de présenter les pratiques habituellement utilisées pour le traitement des données financières ou hydrologiques. Ceci nous a permis de décrire le contexte général et le volet pratique de cette thèse. Dans un deuxième temps, nous procédons à la mise en application des méthodes choisies et à l'interprétation des résultats. Une étape préliminaire de notre application consiste à analyser les caractéristiques des données et à les confronter aux conditions de validité des modèles. Ensuite, dans une seconde partie, il s'agit de présenter les méthodes retenues pour la sélection du seuil optimal, notamment celles qui peuvent être appliquées dans un cadre de non homogénéité (le cas des séries régionales). Deux études de cas seront mises en œuvre, et ce pour différentes méthodes de sélection et d'estimation.

Pour finir, nous résumons les principales conclusions de l'étude effectuée, nous replaçons la contribution de ce travail dans son contexte global de la gestion des risques, et nous proposons quelques perspectives futures.

Avant d'entamer le premier chapitre de notre thèse, nous avons jugés nécessaire, dans cette introduction générale, d'illustrer l'utilisation de la TVE par un recueil d'exemples d'application, tiré du livre de Coles (2001, [23]). Ces exemples, accompagnés par des questions qui sont habituellement soulevées par les praticiens, se réfèrent à plusieurs domaines d'application.

Exemple 1 :

La figure ci-dessous présente les maxima annuels du niveau de la mer au Port Pirie au nord d'Adélaïde, en Australie du Sud au cours de la période 1923-1987.

A partir de ces données, nous pouvons estimer le niveau maximal susceptible de se reproduire dans la région étudiée au cours des 100 ou 1000 ans à venir. Cependant, comment pouvons nous estimer les niveaux qui peuvent se reproduire dans les 1000 prochaines années sans vérifier la validité de la stabilité des changements climatiques durant 1000 ans ?

Certes, les données présentées ne donnent aucun signe sur une non-stationnarité au cours de la période observée, mais cette stabilité peut ne pas persister dans l'avenir, surtout lorsqu'il s'agit d'une période très longue (1000 ans).

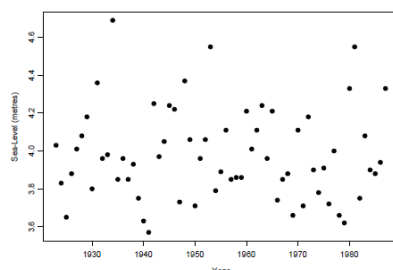


FIGURE 1 – Maxima annuels du niveau de la mer au Port Pirie, Australie du Sud,(Coles 2001, [23]).

Exemple 2 :

Les arguments utilisés pour la modélisation des événements extrêmes fonctionnent aussi bien pour des événements extrêmement petits que pour des événements extrêmement grands. Ceci conduit à un domaine d'application qui concerne la fiabilité.

Conceptuellement, il est souvent raisonnable d'appréhender un grand système comme une composition de nombreuses petites pièces, de telle sorte que l'ensemble du système tombe en panne si l'une des composantes individuelles est en panne.

L'utilisation de la théorie des valeurs extrêmes peut donc contribuer, à détecter le maillon le plus faible, sachant que la fiabilité de l'ensemble du système est égale à celle de la composante la plus faible.

Pour illustrer ce cas, les données ci-dessous présentent la résistance à la rupture de 63 fibres de verre d'une longueur de 1,5 cm, enregistrées dans des conditions expérimentales. Dans ce cas, nous pouvons appréhender une fibre de verre comme un paquet de nombreuses fibres plus fines, de telle sorte que si l'une des petites fibres se casse (les fibres fines), le paquet entier se brise.

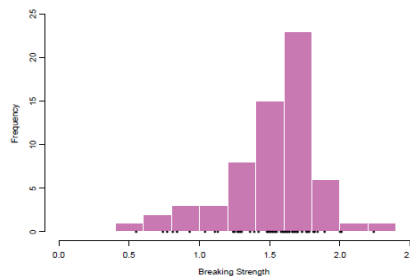


FIGURE 2 – Histogramme de résistance à la rupture de quelques fibres de verre, (Coles 2001, [23]).

Exemple 3 :

L'exemple présent concerne les températures minimales quotidiennes enregistrées en Ohio. La série met en exergue très clairement, un cycle annuel. De ce fait, il ne sera plus raisonnable d'utiliser les modèles des valeurs extrêmes, qui traitent les processus à variation constante au cours du temps.

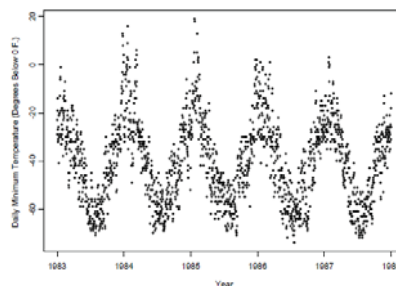


FIGURE 3 – Températures minimales journalières (Ohio), (Coles 2001, [23]).

Exemple 4 :

L'intensité des valeurs extrêmes peut être expliquée par d'autres phénomènes exogènes. A titre d'exemple, les niveaux de pollution sont susceptibles d'être moins extrêmes pendant les périodes de vents violents, ce qui permet de dégager un effet dispersif. Dans les mers australes, ces niveaux peuvent être

exceptionnellement extrêmes pendant les périodes où l'effet El Nino est actif. La figure ci-dessous permet de présenter les maxima annuels des niveaux de la mer dans la région Fremantle, en correspondance avec la valeur annuelle moyenne de l'indice d'oscillation australe (SOI), qui est un proxy pour la volatilité due à des effets météorologiques tel que El Nino. Dans ce cas de figure, les informations supplémentaires par rapport aux maxima annuels peuvent être utilisées pour améliorer l'estimation de la variation des niveaux de la mer au sein de chaque année, ainsi que la variation apparente à travers le temps.

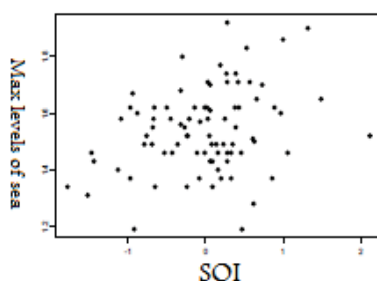


FIGURE 4 – Dépendance apparente de l'indice d'oscillation australe (un indicateur d'El Nino),(Coles 2001, [23]).

Exemple 5 :

La figure ci-dessous présente une série des précipitations journalières cumulées qui correspondent au sud ouest de l'Angleterre, enregistrées au cours de la période allant de 1914 à 1962. Si le modèle statistique approprié pour une telle série était connu, les données présentées peuvent être utilisées pour estimer les paramètres nécessaires. Cependant, en l'absence de cette connaissance, un modèle asymptotique sera préconisé. Pour ce cas, au lieu d'utiliser les blocks des maxima, il est plus judicieux de considérer un événement comme extrême si le niveau des précipitations est au dessus d'un certain seuil (30 mm par exemple).

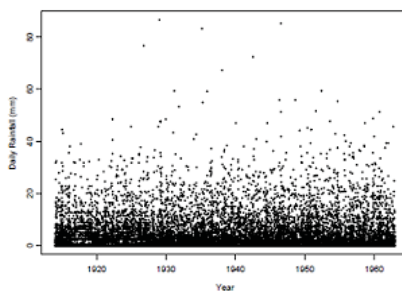


FIGURE 5 – Cumul des précipitations journalières (Sud-ouest de l'Angleterre), (Coles 2001, [23]).

Chapitre 1

Approche des maxima par blocs : lois GEV

1 Valeurs extrêmes et lois α -stables

La loi normale représente l'une des lois de probabilité les plus utilisées pour l'ajustement de très nombreuses variables. Nous pouvons citer à titre d'exemple les rendements agricoles, qui ont constitué la première motivation historique justifiant l'utilisation de la loi normale, les tailles, les poids, les numérations sanguines ou les dosages hormonaux. La loi normale a été introduite par Abraham de Moivre en 1733, comme une loi limite d'une loi binomiale. Elle a été intégrée par la suite, dans les calculs de Gauss en ce qui concerne le mouvement des corps célestes.

L'ample utilisation de la loi normale tient d'une grande part à sa simplicité et d'autre part au théorème central limite (TCL). Ce théorème assure la convergence vers la loi normale, quelle que soit les lois individuelles des échantillons utilisées. Le TCL admet des généralisations qui garantissent la convergence de sommes de variables aléatoires sous des hypothèses faibles.

Malgré sa grande utilité, la loi normale échoue à décrire l'évidence empirique de plusieurs domaines, qui sont caractérisés beaucoup plus par l'existence des extrêmes, telles que les séries financières, hydrologiques ou météorologiques. En finance par exemple, l'hypothèse de normalité des variations des prix est mise à mal par certains mouvements extrêmes des marchés. La loi de probabilité n'est pas exactement normale mais plus large sur les bords et plus cintrée. Pire encore, les prix des obligations sont compliqués pour être modélisés par une loi normale, car ils dépendent des différents taux d'intérêt, eux-mêmes très liés entre eux par une courbe des taux aux déformations limitées (Mélou,

(2008,[99])). Par ailleurs, les lois de probabilité en hydrologie fréquentielle ne sont pas symétriques, ce qui constitue un obstacle à l'utilisation de la loi normale.

Pour pallier ces problèmes et mieux approcher les extrêmes, les statisticiens ont eu tendance à utiliser les lois stables (Lévy (1924,[85]), Fama (1965,[47]), Mandelbrot(1963,[97])). Ces lois plus flexibles que la loi normale, présentent plusieurs propriétés attractives pour les praticiens. Elles permettent de rendre compte des queues lourdes et du comportement asymétrique.

Les définitions suivantes ont été proposées pour caractériser les lois stables :

Définition 1.1 *On dit qu'une variable aléatoire X suit une loi **stable** si pour tout $n \geq 2$, il existe un réel strictement positif C_n et un réel D_n tel que :*

$$X_1 + X_2 + \dots + X_n \stackrel{loi}{=} C_n X + D_n \quad (1.1)$$

Où X_1, X_2, \dots, X_n sont n copies indépendantes de X .

Le réel strictement positif C_n est nécessairement de la forme $n^{1/\alpha}$, pour un certain $\alpha \in]0, 2]$, d'où l'appellation loi α -stable (Nikias et Shao(1995,[102])).

Définition 1.2 *On dit qu'une variable aléatoire X suit une loi α -stable si et seulement si sa fonction caractéristique est de la forme :*

$$\phi(t) = \exp\{iat - \gamma|t|^\alpha[1 + i\beta\text{sign}(t)\omega(t, \alpha)]\} \quad (1.2)$$

Avec :

$$\omega(t, \alpha) \begin{cases} -\tan\frac{\alpha\pi}{2} & \text{si } \alpha \neq 1 \\ \frac{2}{\pi} \log|t| & \text{si } \alpha = 1 \end{cases} \quad (1.3)$$

Où $\alpha \in]0, 2]$, $\beta \in]-1, 1[$, $\gamma > 0$ et $a \in \mathbb{R}$. Les paramètres α , a , γ et β représentent respectivement l'exposant caractéristique de la loi stable, le paramètre de position, le paramètre d'échelle et le paramètre d'asymétrie.

Le théorème suivant montre que les lois stables peuvent être utilisées particulièrement, pour approximer la loi d'une somme de variables aléatoire i.i.d.

Théorème 1.3 ([103]) *Soit X_1, X_n, \dots des variables aléatoires indépendantes, de même loi. Une variable aléatoire X est limite en loi de la suite*

$$\frac{X_1 + \dots + X_n}{a_n} - b_n$$

pour une suite de réels strictement positifs (a_n) et une suite de réels (b_n) , si et seulement si, X suit une loi stable.

(Pour la démonstration de ce théorème, voir (Nolan(1996,[103]))).

Les lois α -stables présentent un grand intérêt dans la modélisation de nombreux problèmes physiques. La caractéristique de ces lois est leur index de stabilité α qui indique la vitesse de décroissance des queues. Ainsi, les lois peuvent être regroupées selon le comportement de leurs queues. La classification suivante a été proposée récemment par El Adlouni et al.(2007,[41]) :

- (E) : Les lois exponentielles dont les moments n’existent pas.
- (D) : Les lois subexponentielles.
- (C) : Les lois à variations régulières.
- (B) : Les lois avec un comportement de Pareto.
- (A) : Les lois α -stables avec $\alpha < 2$.

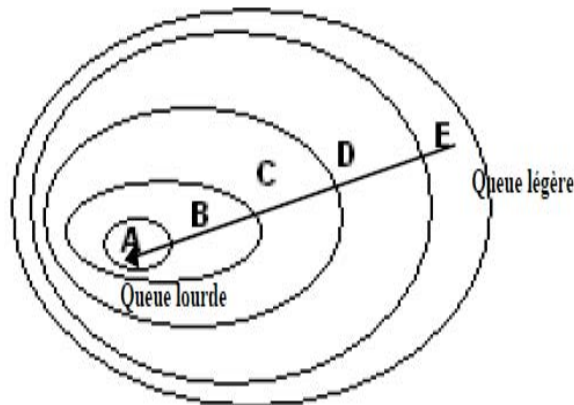


FIGURE 1.1 – Classification des lois selon les queues, (El Adlouni et al.(2007,[41]))

La classe (E) contient toutes les lois tel que $\mathbb{E}[e^X] = \infty$. La probabilité au dépassement, $\bar{F} = Pr(X \geq x) = 1 - F(x)$, pour les extrêmes de cette classe, décroît moins rapidement que celle de la loi normale. La classe (E) est donc celle des lois ayant une queue plus lourde que la loi normale.

La classe (D) quant à elle, contient les lois tel que \bar{F} décroît plus lentement que n’importe quelle loi exponentielle.

Pour la classe (C) dite classe des lois à variations régulières, la probabilité au dépassement des extrêmes décroît suivant une fonction puissance (appelée aussi décroissance géométrique).

La classe (B) est celle des lois de Pareto. Enfin, la classe (A) est celle qui

regroupe des lois à différentes asymétries, avec des queues très lourdes.

2 Théorème fondamental des valeurs extrêmes

Le résultat de base de la TVE consiste à décrire la loi asymptotique du maximum de n variables indépendantes et identiquement distribuées (i.i.d). La forme de cette loi (notée GEV) dépend d'un seul paramètre qui permet de spécifier le comportement de la queue de la loi considérée. Ce paramètre est l'indice des valeurs extrêmes (noté ξ). En d'autres termes, le résultat de base de la TVE assure que la loi du maximum de n variables (i.i.d) est toujours dans le domaine d'attraction d'une loi GEV_ξ .

Définition 2.1 Soit G une fonction de répartition définie sur \mathbb{R} . Une fonction de répartition F sera dite appartenir au domaine d'attraction $\mathcal{D}(G)$ de G s'il existe deux suites (a_n) ($a_n > 0$) et (b_n) tel que :

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x) \quad (2.4)$$

pour tout point de continuité x , de G .

Dans le cas où F est connue, G peut être identifiée à partir de la loi des statistiques d'ordre. En effet, si nous supposons que les données observées sont des réalisations d'une suite de variables aléatoires X_1, X_2, \dots, X_n (i.i.d) de loi F , alors la fonction de répartition de la statistique d'ordre $X_{i:n}$ est donnée par :

$$F_{i:n}(x) = P(X_{i:n} \leq x) = \sum_{k=1}^n C_n^k [F(x)]^k [1 - F(x)]^{n-k} \quad (2.5)$$

En conséquence, les lois de $M_n = X_{n,n} = \max(X_1, X_2, \dots, X_n)$ et de $W_n = X_{1,n} = \min(X_1, X_2, \dots, X_n)$, notées respectivement F_M et F_W , sont données par :

$$\forall x \in \mathbb{R} : F_M(x) = [F(x)]^n; F_W(x) = 1 - [1 - F(x)]^n \quad (2.6)$$

Dans la plupart des cas F n'est pas connue et parfois, l'utilisation de la loi des statistiques d'ordre aboutissent à des calculs compliqués. D'où l'utilité d'établir un résultat asymptotique (résultat de base de la TVE), connu sous le nom du théorème de Fisher-Tippett (1928).

Par analogie avec le théorème central limite (TCL), le théorème de Fisher-Tippett établit la convergence en loi de M_n (le théorème est aussi applicable à W_n) :

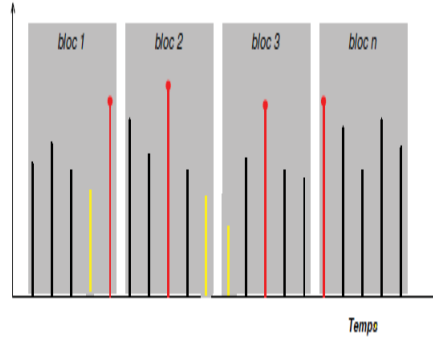


FIGURE 1.2 – Méthode des maxima (minima) par bloc.

Théorème 2.2 (Fisher-Tippett) ([44]). Soient $X_1, \dots, X_i, \dots, X_n$ n variables aléatoires indépendantes et de même loi de probabilité F . S'il existe des constantes a_n et b_n et une loi G non dégénérée (c'est à dire différente d'une masse de Dirac) tel que :

$$\lim_{n \rightarrow \infty} Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = G(x), \forall x \in \mathbb{R}$$

Alors G appartient à l'un des trois types de loi :

Loi de Gumbel ($\xi = 0$) :

$$\Lambda_{\mu, \sigma}(x) = \exp\left(-\exp\left(-\frac{x - \mu}{\sigma}\right)\right), x \in \mathbb{R} \quad (2.7)$$

Loi de Fréchet ($\xi > 0$) :

$$\Phi_{\mu, \sigma, \xi}(x) = \begin{cases} \exp\left(-\left(\frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) & \text{si } x > \mu \\ 0 & \text{si } x \leq \mu \end{cases} \quad (2.8)$$

Loi de Weibull ($\xi < 0$) :

$$\Psi_{\mu, \sigma, \xi}(x) = \begin{cases} \exp\left(-\left(-\frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) & \text{si } x < \mu \\ 1 & \text{si } x \geq \mu \end{cases} \quad (2.9)$$

Les trois lois de probabilité ci-dessus sont appelées lois des valeurs extrêmes.

Pour la preuve, nous renvoyons à Embrechts et al. (1997, [44]), chapitre 3, p. 122. Il faut bien signaler que chacune des trois lois des valeurs extrêmes peut s'obtenir par une transformation fonctionnelle de l'autre. D'une façon similaire, on définit les lois des valeurs extrêmes associées au minimum.

([44]) S'il existe des suites de normalisation $a_n > 0$ et $b_n \in \mathbb{R}$ et une loi non dégénérée H^* tel que :

$$\lim_{n \rightarrow \infty} P\left[\frac{W_n - b_n}{a_n} \leq x\right] = \lim_{n \rightarrow \infty} (1 - F(a_n x + b_n))^n = H^*(x) \quad (2.10)$$

alors H^* est une loi des valeurs extrêmes associée au minimum.

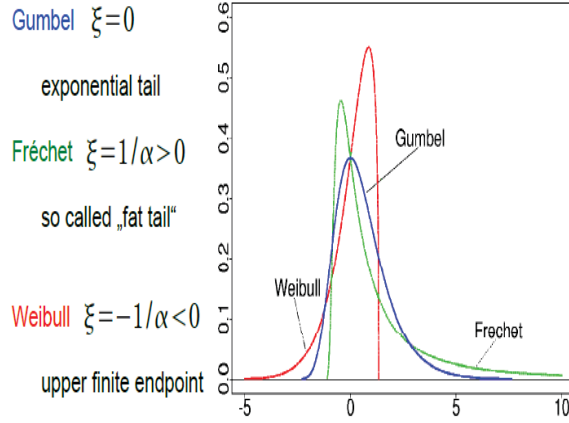


FIGURE 1.3 – Lois des valeurs extrêmes.

En effet : $\min(X_i) = -\max(-X_i)$. Ainsi, $\forall x \in \mathbb{R}, \forall 1 \leq i \leq n$:

$$P[\min(X_i) > x] = P[\max(-X_i) \leq -x] \Rightarrow 1 - F_W(x) = F_M(-x)$$

Von Mises (1954, [146]) et Jenkinson (1955, [71]) ont établi un résultat théorique qui permet d’identifier une famille de lois GEV, qui regroupe les trois types de lois associés aux valeurs extrêmes. Le résultat théorique est formulé par le biais du théorème suivant :

Théorème 2.4 ([117]) *S’il existe des suites de normalisation $a_n > 0$ et $b_n \in \mathbb{R}$ et une loi G^* tel que :*

$$\lim_{n \rightarrow \infty} P\left[\frac{M_n - b_n}{a_n} \leq x\right] = G^*(x); \forall x \in \mathbb{R} \tag{2.11}$$

alors G^* est donnée par :

$$G_{\mu, \sigma, \xi}(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \tag{2.12}$$

Où x est tel que $1 + \xi\left(\frac{x - \mu}{\sigma}\right) > 0$, $-\infty < \mu < +\infty$, $\sigma > 0$ et $-\infty < \xi < +\infty$. Les paramètres μ , σ et ξ sont respectivement, les paramètres de position, d’échelle et de forme.

Une démonstration détaillée de ce théorème est donnée dans l’ouvrage de Resnick (1987, [117]), avec des développements dans le livre d’Embrechts et al. (1997, p. 152, [44]) et des illustrations dans le livre de Galambos (1987, pp.53-54, [49]).

Il faut bien noter que le paramètre ξ conditionne le type de la loi des valeurs extrêmes. Nous présentons dans ce qui suit, les domaines d’attraction dans les trois cas correspondant au paramètre ξ .

1. Cas où $\xi = 0$:

La loi présente dans la queue une décroissance de type exponentiel, ce qui permet de caractériser dans ce cas, le domaine d'attraction de Gumbel $\mathcal{D}(\Lambda)$. Ce dernier est délicatement traitable, car il n'y a pas de lien direct entre la queue de la loi et les fonctions à variations lentes définies par (Delmas et Jourdain (2006, [31])) :

Définition 2.5 Une fonction L est dite à variation lente si $L(t) > 0$ pour t assez grand et si pour tout $x > 0$, on a :

$$\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1 \quad (2.13)$$

Von Mises (1936, [145]) a donné une caractérisation simple pour le domaine d'attraction de Gumbel, formulée par le biais du théorème suivant :

Théorème 2.6 ([145]) S'il existe une fonction mesurable R , appelée fonction auxiliaire tel que :

$$\lim_{x \rightarrow \omega(F)} \frac{1 - F(t + xR(t))}{1 - F(t)} = \exp(-x) \quad (2.14)$$

avec $\omega(F) = \sup\{x \in \mathbb{R} : F(x) < 1\}$ est le point terminal de F , alors $F \in \mathcal{D}(\Lambda)$.

Le tableau suivant fournit quelques exemples de lois qui appartiennent au domaine d'attraction de Gumbel.

Tab. 2.1 : Lois du domaine d'attraction de Gumbel

Loi	$1 - F(x)$
Benktander $II, \alpha, \beta > 0$	$x^{-(1-\beta)} \exp(-\frac{\alpha}{\beta} x^\beta)$
Logistic	$\frac{1}{1+\exp(x)}$

2. Cas où $\xi < 0$:

Le domaine d'attraction dans ce cas est celui de Weibull $\mathcal{D}(\Psi_\xi)$. Les lois de ce domaine sont bornées à droite, et par conséquent, le point terminal $\omega(F)$ est fini. Une caractérisation d'appartenance à ce domaine d'attraction est donnée par le théorème suivant (pour la démonstration, voir Gnedenko (1943, [54])) :

Théorème 2.7 ([54]) Une fonction de répartition F appartient au $\mathcal{D}(\Psi_\xi)$ si et seulement si $\omega(F) < +\infty$ et

$$\overline{F}(\omega(F) - \frac{1}{x}) = x^{-\frac{1}{\xi}} l(x) \quad (2.15)$$

avec : \overline{F} est la fonction de survie donnée par $\overline{F}(x) = 1 - F(x)$, l est une fonction à variation lente.

Le tableau ci-dessous présente quelques lois qui appartiennent au domaine d'attraction de Weibull.

Tab. 2.2 : Lois du domaine d'attraction de Weibull

Loi	$1 - F(\omega(F) - \frac{1}{x})$	ξ	$l(x)$
Uniforme	$\frac{1}{x}, x > 1$	-1	1
Weibull	$1 - \exp(-x^{-\alpha})$	$-\frac{1}{\alpha}$	$1 - \frac{x^{-\alpha}}{2} + o(x^{-\alpha})$
Reverse Burr (λ, β, τ)	$(\frac{\beta}{\beta+x^\tau})^\lambda$ avec $x > 0$	$-\frac{1}{\lambda\tau}$	$\beta^\lambda(1 - \lambda\beta x^{-\tau} + o(x^{-\tau}))$

3. **Cas où $\xi > 0$:**

Ce cas correspond au domaine d'attraction de Fréchet. Les lois appartenant à ce domaine d'attraction sont caractérisées par une queue à décroissance lente (polynomiale) à l'infini, et un point terminal $\omega(F) = +\infty$. Elles sont dites aussi lois à queues lourdes. Une caractérisation de ce domaine d'attraction noté $\mathcal{D}(\Phi_\xi)$ est donnée par le théorème suivant (pour la démonstration, voir Gnedenko (1943, [54])) :

Théorème 2.8 ([54]) *Une fonction de répartition F appartient au $\mathcal{D}(\Phi_\xi)$ si et seulement si sa fonction de survie est donnée par :*

$$\bar{F}(x) = x^{-\frac{1}{\xi}} l(x) \tag{2.16}$$

où l est une fonction à variation lente.

Le tableau ci-dessous présente quelques lois qui appartiennent au domaine d'attraction de Fréchet.

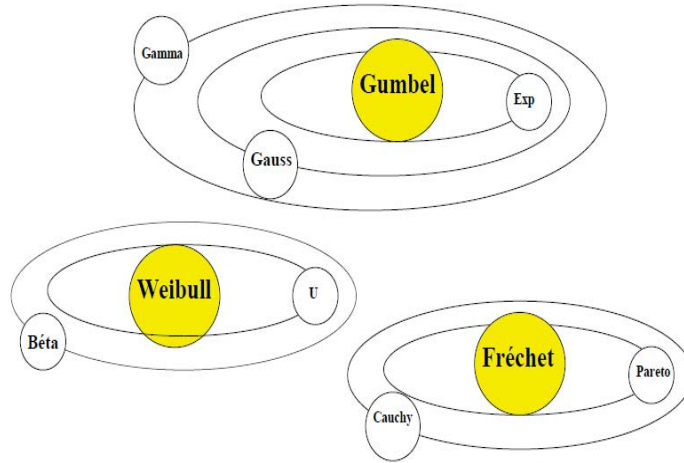
Tab. 2.3 : Lois du domaine d'attraction de Fréchet

Loi	$1 - F(x)$	ξ	$l(x)$
Pareto (α) avec $\alpha > 0$	$x^{-\alpha}$	$\frac{1}{\alpha}$	1
Burr (η, τ, λ) avec $\eta, \tau, \lambda > 0$	$(\frac{\eta}{\eta+x^\tau})^\lambda, x > 0$	$\frac{1}{\lambda\tau}$	$(\frac{\eta}{1+\frac{\eta}{x^\tau}})^\lambda$
Pareto Généralisé (σ, ξ) avec $\sigma, \xi > 0$	$(1 + \frac{\xi x}{\sigma})^{-\frac{1}{\xi}}, x > 0$	ξ	$(\frac{\sigma}{\xi})^{\frac{1}{\xi}} (1 + \frac{\sigma}{\xi x})^{-\frac{1}{\xi}}$

Remarques :

1. Une fonction de répartition F peut appartenir à deux domaines d'attraction différents \mathcal{D}_1 et \mathcal{D}_2 (pour plus de détails, voir Resnick (1987, [117])).
2. Les variables aléatoires dans les trois domaines d'attraction de Gumbel, de Fréchet et de Weibull sont liées par la relation suivante (voir Embrecht et al. (1997, [44])) :

$$X \in \mathcal{D}(\Phi_\alpha) \Leftrightarrow \log(X^{\frac{1}{\xi}}) \in \mathcal{D}(\Lambda) \Leftrightarrow -X^{-1} \in \mathcal{D}(\Psi_\xi) \tag{2.17}$$

FIGURE 1.4 – Lois du domaine d’attraction $\mathcal{D}(\Lambda)$, $\mathcal{D}(\Phi_\xi)$ ou $\mathcal{D}(\Psi_\xi)$.

3. Les trois domaines d’attraction $\mathcal{D}(\Lambda)$, $\mathcal{D}(\Phi_\xi)$ et $\mathcal{D}(\Psi_\xi)$ ne sont pas restrictifs. Il se peut qu’une loi n’appartienne à aucun de ces trois domaines d’attraction ; c’est le cas de la loi définie par :

$$F(x) = 1 - \frac{1}{\log(x)}, x \geq e. \quad (2.18)$$

Outre les domaines d’attraction qui caractérisent chacune des lois GEV, les champs d’application conditionnent leurs choix. En effet, la loi de Gumbel par exemple, sert de référence car elle permet de mesurer l’écart entre les lois empiriques et la loi normale. Elle est généralement utilisée pour prévoir le niveau des crues d’un fleuve. Elle peut aussi servir à prédire la probabilité d’un événement critique comme un tremblement de terre.

En ce qui concerne la loi de Fréchet, elle est couramment évoquée en finance et en macroéconomie, où les variations sont généralement non bornées. Par ailleurs, cette loi s’applique en hydrologie pour modéliser des événements extrêmes tels que les débits maximaux des rivières.

La loi de Weibull quant à elle, est largement utilisée grâce à sa grande flexibilité. Elle est préconisée en particulier, dans les analyses de durée de vie ou pour l’estimation des potentiels éoliens.

3 Estimation des modèles GEV

Les lois GEV se distinguent essentiellement selon le paramètre ξ . Ainsi, l’estimation sera établie selon deux cas : $\xi = 0$ (loi de Gumbel) et $\xi \neq 0$ (lois

de Fréchet et de Weibull).

La littérature propose une multitude de méthodes, consacrées principalement à l'estimation des paramètres des lois GEV. Nous pouvons citer par exemple, les méthodes d'estimation empirique (Gumbel et Mustafi (1967, [59]) ou Tiago de Oliveira (1975, [138])), la méthode du maximum de vraisemblance (Smith (1987, [131] et 1985, [132]), Prescott et Walden (1980, [111] et 1983, [110]), Tiago de Oliveira (1982, [139]), Hougaard (1986, [64])), la méthode des moments (Christopeit, (1994, [22])), la méthode des moments de probabilité pondérés (Greenwood et al., (1979, [51])) ou encore, les méthodes bayésiennes (Lye et al., (1993, [95])).

La comparaison de ces méthodes d'estimation s'effectue dans la plupart du temps dans un cadre empirique (voir Benkhaled (2007,[5]) et Lang et al. (2012, [82])).

Par ailleurs, d'autres approches principalement non paramétriques, ont été dédiées à l'estimation de l'indice de queue. Nous pouvons citer à titre d'exemple l'estimateur de Pickands (1975, [107]), l'estimateur de Hill (1975, [60]) (pour la loi de Fréchet uniquement) et l'estimateur de Dekkers-Einmahl-de Hann (Dekkers et al. (1989, [38])). Ces approches non paramétriques seront détaillées au chapitre suivant.

3.1 Estimation pour la loi de Gumbel

Dans cette partie, nous notons la loi de Gumbel par $G_0(\mu_0, \sigma_0)$, où μ_0 et σ_0 désignent respectivement, les paramètres de position et d'échelle ($\xi = 0$ dans le cas de la loi de Gumbel). La fonction de densité de Gumbel est notée g_0 .

4.1.1 Méthode de la régression linéaire :

Soient Y_1, \dots, Y_m les maxima par bloc issus d'une loi de Gumbel et soient $Y_{1,m} \leq \dots \leq Y_{m,m}$ les statistiques d'ordre associées à Y_1, \dots, Y_m .

La méthode d'estimation par régression linéaire consiste à ajuster le nuage de points :

$$(-\log(-\log(G_0(Y_{i,m}))); Y_{i,m}) \quad (3.19)$$

à une droite d'équation $Y_{i,m} = aX + b$, avec $X = -\log(-\log(G_0(Y_{i,m})))$, en calculant les coefficients a et b par la méthode des moindres carrés.

Le coefficient directeur (a) et l'ordonnée à l'origine (b) de la droite ajustée,

permettent de calculer les estimateurs $\widehat{\mu}_0^{reg}$ et $\widehat{\sigma}_0^{reg}$:

$$\begin{cases} \widehat{\mu}_0^{reg} = b \\ \widehat{\sigma}_0^{reg} = a \end{cases} \quad (3.20)$$

L'estimation des quantiles extrêmes dépassés avec une faible probabilité p , pour la loi de Gumbel G_0 et pour une période de retour $T = \frac{1}{p} = \frac{1}{1-G_0(y)}$, est obtenue en remplaçant μ et σ par leurs estimations dans la formule suivante :

$$q_0 = \mu_0 + \sigma_0 \left(-\log \left(-\log \left(1 - \frac{1}{T} \right) \right) \right) \quad (3.21)$$

Ainsi, l'estimateur du quantile extrême \widehat{q}_0^{reg} est :

$$\widehat{q}_0^{reg} = \widehat{\mu}_0^{reg} + \widehat{\sigma}_0^{reg} \left(-\log \left(-\log \left(1 - \frac{1}{T} \right) \right) \right) \quad (3.22)$$

4.1.2 Méthode du maximum de vraisemblance :

Cette méthode consiste à chercher les paramètres $\Theta_0 = (\mu_0, \sigma_0)$ qui maximisent la fonction de vraisemblance $l(\mu_0, \sigma_0) = \prod_{i=1}^m g_0(y_i, \mu_0, \sigma_0)$ ou logvraisemblance $L(\mu_0, \sigma_0) = \log l(\mu_0, \sigma_0)$, en résolvant le système suivant :

$$\begin{cases} \frac{\partial \log L(\Theta_0)}{\partial \mu_0} = 0 \\ \frac{\partial \log L(\Theta_0)}{\partial \sigma_0} = 0 \end{cases} \quad (3.23)$$

Cependant, la solution n'est pas explicite et il faut passer par une résolution numérique en utilisant par exemple, la méthode de *Newton – Raphson*.

4.1.3 Méthode des moments :

La méthode consiste à égaliser respectivement, les deux premiers moments théoriques m_1 et m_2 aux deux premiers moments observés : moyenne (\bar{y}_m) et variance S_Y^2 empiriques .

Dans le cas d'une variable aléatoire issue d'une loi de Gumbel, la méthode du maximum de vraisemblance abouti au système suivant :

$$\begin{cases} \mathbb{E}[Y] = \mu_0 + \gamma_* \sigma_0 = \bar{y}_m \\ \mathbb{V}[Y] = \frac{1}{6} \pi^2 \sigma_0^2 = S_Y^2 \end{cases} \quad (3.24)$$

où $\gamma_* \approx 0,57721$ représente la constante d'*Euler*. La résolution de ce système d'équation fournit les valeurs des estimateurs $\widehat{\mu}_0^{mom}$ et $\widehat{\sigma}_0^{mom}$:

$$\begin{cases} \widehat{\sigma}_0^{mom} = \frac{\sqrt{6} S_Y}{\pi} \\ \widehat{\mu}_0^{mom} = \bar{y}_m - \gamma_* \widehat{\sigma}_0^{mom} \end{cases} \quad (3.25)$$

3.2 Estimation pour les lois de Fréchet et de Weibull

Pour $\xi \neq 0$, les domaines d'attraction sont ceux de Fréchet et de Weibull. Ces lois seront notées $G_\xi(\mu, \sigma, \xi)$ ($\xi > 0$ pour une loi de Fréchet et $\xi < 0$ pour une loi de Weibull). La fonction de densité sera notée g_ξ .

Nous présentons pour les deux lois de Fréchet et de Weibull, les méthodes d'estimation du maximum de vraisemblance et des moments de probabilité pondérés.

4.2.1 Méthode du maximum de vraisemblance :

La méthode consiste à chercher les paramètres $\Theta = (\mu, \sigma, \xi)$ qui maximisent la fonction de log-vraisemblance. Pour l'obtention de $\hat{\mu}^{mv}$, $\hat{\sigma}^{mv}$ et $\hat{\xi}^{mv}$, la maximisation conduit à résoudre le système suivant :

$$\begin{cases} \frac{\partial \log L(\Theta)}{\partial \sigma} = 0 \\ \frac{\partial \log L(\Theta)}{\partial \xi} = 0 \\ \frac{\partial \log L(\Theta)}{\partial \mu} = 0 \end{cases} \quad (3.26)$$

Smith (1985, [132]) a étudié le problème de l'estimation par la méthode du maximum de vraisemblance. Il a obtenu les résultats suivants :

- Si $\xi > -0.5$, les estimateurs du maximum de vraisemblance possèdent des propriétés asymptotiques telle que la convergence vers la vraie valeur du paramètre inconnu, l'invariance par rapport à une transformation paramétrique et l'efficacité asymptotique.
- Si $-1 < \xi < -0.5$, les estimateurs du maximum de vraisemblance ne possèdent pas les propriétés asymptotiques standards.
- Si $\xi < -1$, l'obtention des estimateurs du maximum de vraisemblance n'est pas garantie.

4.2.2 Méthode des moments de probabilité pondérés PWM :

Cette méthode a été introduite par Greenwood et al. (1979, [51]) dans le but de surmonter les problèmes de convergence, surtout pour les petits échantillons ($n < 25$) (Lubes et al. (1991, [94])).

Les moments de probabilité pondérés sont définis par :

$$M_{p,r,s} = \mathbb{E}[Y^p (F(Y))^r (1 - F(Y))^s] \quad (3.27)$$

où : p, r, s sont des réels et Y est une variable aléatoire issue de la loi F .

Dans le cas où $\xi \neq 0$, un choix possible pour calculer les moments $M_{p,r,s}$ est de prendre $p = 1$, $s = 0$ et $r = 0, 1, 2$ (Beirlant et al. (2004, [2])). Ainsi :

$$M_{1,r,0} = \frac{1}{r+1} \left\{ \mu - \frac{\sigma}{\xi} [1 - (r+1)^\xi \Gamma(1-\xi)] \right\}, \quad \xi < 1 \quad (3.28)$$

Les estimateurs *PWM* déduits de la méthode des moments de probabilité pondérés sont notés $\widehat{\mu}^{pwm}$, $\widehat{\sigma}^{pwm}$ et $\widehat{\xi}^{pwm}$. Ils sont obtenus suite à la résolution d'un système d'équations, pour $r = 0, 1, 2$:

$$\begin{cases} M_{1,0,0} = \mu - \frac{\sigma}{\xi} (1 - \Gamma(1-\xi)) \\ 2M_{1,1,0} - M_{1,0,0} = \frac{\sigma}{\xi} \Gamma(1-\xi) (2^\xi - 1) \\ \frac{3M_{1,2,0} - M_{1,0,0}}{2M_{1,1,0} - M_{1,0,0}} = \frac{3^\xi - 1}{2^\xi - 1} \end{cases} \quad (3.29)$$

Dans la pratique, Hosking et al. (1985, [63]) préconisent l'utilisation d'un estimateur asymptotiquement consistant, donné par :

$$\widehat{M}_{1,r,0} = \frac{1}{m} \sum_{i=1}^m (\widehat{F}(Y_{i,m}))^r Y_{j,m} \quad (3.30)$$

Avec :

$$\widehat{F}(y_{i,m}) = \frac{i-a}{m}, \quad 0 < a < 1 \quad (3.31)$$

$$\widehat{F}(y_{i,m}) = \frac{i-a}{m+1-2a}, \quad -\frac{1}{2} < a < \frac{1}{2} \quad (3.32)$$

4 Non stationnarité et dépendance

La loi des extrêmes pour une suite de variables aléatoires i.i.d a été élaborée à partir d'une normalisation. Cette approche permet de se retrouver dans le contexte de l'application du théorème central limite. Cependant, des extensions de la théorie classique ont rapidement vu le jour, suite à la prise de conscience concernant les limites de l'approche traditionnelle. Ces limites sont liées aux hypothèses assez lourdes d'indépendance et de stationnarité. Ces hypothèses s'imposent comme exigence souvent mise en défaut, notamment en finance ou en hydrologie.

La dépendance temporelle marque plusieurs événements extrêmes qui affichent généralement, une dépendance intrinsèque en raison de leurs autocorrélations ou de l'effet des variables exogènes. En plus, les conditions extrêmes persistent souvent pour certaines observations consécutives. A titre d'exemple, nous pouvons intuitivement penser qu'il est plus probable que la pluie à une certaine

heure soit extrême, si celle de l'heure précédente l'était, et souvent suite à une tempête forte, les débits maximaux d'une rivière se produisent ensemble.

Pour remédier à cette situation, nous pouvons traiter dans une première étape, le cas des observations identiquement distribuées mais dépendantes, formant par conséquent un processus stationnaire (Rosenblatt (1956, [124]) ou Leadbetter, Lindgren et Rootzèn (1983, [83])). Ainsi, l'une des premières pistes naturelles de généralisation de la théorie traditionnelle des valeurs extrêmes, consiste à remettre en cause l'hypothèse de dépendance des variables aléatoires traitées. Les travaux initiaux dans ce domaine sont ceux de Watson (1954, [147]), Loynes (1965, [93]) et Welsch (1971, [148]). Ces travaux s'intéressaient à l'analyse des suites de processus fortement dépendants.

L'étude des suites de processus gaussiens pour lesquels des restrictions simples ont été imposées aux autocorrélations (dépendance par exemple), figuraient dans les recherches de Berman (1971, [6]). De son côté et sous des hypothèses relatives aux autocorrélations, Berman (1964, [7]) a montré que la loi des maxima d'une suite stationnaire (X_n) s'identifie à une loi GEV classique (le même type de loi que celui des maxima d'une suite de variable i.i.d). Par ailleurs, le cas des variables aléatoires m -dépendantes a été traité par Lai et Robbins (1976, [81]), Philipp (1971, [106]), Ibrajimov et Linnik (1971, [69]). Les problèmes de dépendance ont été aussi abordés dans le cas des processus bivariés (Pickands (1981, [108]), Deheuvels (1984, [36]), Tawn (1988, [134])) ou plus généralement dans un cadre multivarié (Resnick (1987, [118]) ou Hüsler (1994, [67])).

Dans une catégorie à part, nous pouvons regrouper les travaux tels que ceux proposés par Hüsler (1985, [68]). Ces recherches remettent en cause à la fois l'hypothèse de stationnarité et celle d'indépendance dans l'analyse des valeurs extrêmes. Cependant, cette dernière catégorie d'étude demeure peu développée.

4.1 Les extrêmes dans le cas des suites d'observations dépendantes et stationnaires :

La théorie des extrêmes dans le cas d'observations dépendantes, consiste à proposer des réponses appropriées aux mêmes questions que dans le cas classique, mais dans une situation de dépendance qui répond à la réalité de la plupart des séries traitées, tout en exigeant l'hypothèse de la stationnarité. La stationnarité caractérise un processus pour lequel les propriétés stochastiques sont homogènes à travers le temps (sans forcément que les variables aléatoires

sous-jacentes soient mutuellement indépendantes). Dorénavant, nous ne traitons que les processus stationnaires. En termes précis, nous pouvons définir la stationnarité comme suit :

Définition 4.1 : (*Stationnarité forte*). Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires. On dit que $(X_n)_{n \in \mathbb{N}}$ est stationnaire si pour tout $t \in \mathbb{N}$, et pour tout $k \in \mathbb{N}$, la loi conjointe de (X_1, \dots, X_t) est égale à la loi conjointe de $(X_{1+k}, \dots, X_{t+k})$.

Le théorème qui généralise le cas classique de la TVE aux processus stationnaires, nécessite la validation d'une condition d'application notée $D(u_n)$ (Leadbetter, Lindgren et Rootzèn (1983, [83])) :

Condition $D(u_n)$ ([83]) : Etant donné une suite de réels (u_n) et une suite d'entiers naturels $i_1 < i_2 < \dots < i_p < j_1 < j_2 < \dots < j_q \leq n$, tel que $j_1 - i_p \geq \ell$, la condition $D(u_n)$ est vérifiée si on a :

$$|F_{i_1, \dots, i_p, j_1, \dots, j_q}(u_n) - F_{i_1, \dots, i_p}(u_n)F_{j_1, \dots, j_q}(u_n)| \leq \alpha_{n, \ell_n}$$

Où α_{n, ℓ_n} ne croît pas avec ℓ_n , $\lim_{n \rightarrow \infty} \alpha_{n, \ell_n} = 0$ et $\frac{\ell_n}{n} \rightarrow 0$.

Cette condition traduit le fait qu'en général, la corrélation temporelle des observations décroît, si la période qui les sépare augmente. Dans la pratique, Cette condition permet de ne considérer que les processus stationnaires dont la dépendance entre observations extrêmes ne se manifeste qu'à court terme. Quant à la dépendance à long terme, elle peut être négligée. Le théorème classique de la TVE peut être généralisé comme suit :

Théorème 4.2 (*Leadbetter*) ([83]). Soient X_1, \dots, X_n une suite stationnaire et deux suites réelles (c_n) dans \mathbb{R}^{+*} et (d_n) dans \mathbb{R} tel que :

$$P(c_n^{-1}(M_n - d_n) \leq x) \rightarrow G_\xi(x) \quad (4.33)$$

Si la condition $D(u_n)$ est vérifiée pour $u_n = xc_n + d_n$, alors la loi $G_\xi(x)$ est de même type que l'une des trois loi suivantes :

Loi de Gumbel ($\xi = 0$) :

$$\Lambda_{\mu, \sigma}(x) = \exp\left(-\exp\left(-\frac{x - \mu}{\sigma}\right)\right), x \in \mathbb{R} \quad (4.34)$$

Loi de Fréchet $\xi > 0$

$$\Phi_{\mu, \sigma, \xi}(x) = \begin{cases} \exp\left(-\left(\frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) & x > \mu \\ 0 & x \leq \mu \end{cases} \quad (4.35)$$

Loi de Weibull ($\xi < 0$) :

$$\Psi_{\mu,\sigma,\xi}(x) = \begin{cases} \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{\frac{1}{\xi}}\right) & \text{si } x < \mu \\ 1 & \text{si } x \geq \mu \end{cases} \quad (4.36)$$

Ce résultat généralise le théorème classique de la TVE dans le sens où il implique que les maxima issus d'une série stationnaire à observations dépendantes, suivent une loi GEV. Ainsi, si les données traitées sont stationnaires et si la corrélation entre les observations semble décroître avec le décalage temporel qui les sépare, alors, par le biais du théorème 4.2, la loi *GEV* reste une approximation valable pour ajuster les données observées.

4.2 GEV non stationnaires :

Dans le cas des processus non stationnaires, les paramètres ne sont pas constants, mais ils dépendent du temps ou d'autres covariables d'une façon linéaire ou sous forme fonctionnelle plus générale.

Nous présentons dans ce qui suit le cas de dépendance linéaire par rapport au temps, vu que les autres formes de dépendance peuvent être ramenées au cas linéaire. Les covariables, autres que le temps, se traitent de la même façon.

Soit $U = (U_1, U_2, \dots, U_{n_\mu})'$ le vecteur de n_μ covariables du paramètre de position μ ,

$$\mu_t = U'(t)\beta = \sum_{i=1}^{n_\mu} \beta_i U_i(t) \quad (4.37)$$

où $(\beta_1, \beta_2, \dots, \beta_{n_\mu})'$ est le vecteur des paramètres caractérisant la dépendance de μ_t avec le vecteur des covariables U .

Soit $V = (V_1, V_2, \dots, V_{n_\sigma})'$ le vecteur des covariables associées à σ_t , on a :

$$\phi_t = \log(\sigma_t) = V' \cdot \delta = \sum_{i=1}^{n_\sigma} \delta_i V_i(t) \quad (4.38)$$

De la même façon, le paramètre de forme peut s'écrire en fonction d'un vecteur de covariables $W = (W_1, W_2, \dots, W_{n_\xi})'$ et un vecteur des paramètres $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{n_\xi})'$:

$$\xi_t = W'(t) \cdot \gamma = \sum_{i=1}^{n_\xi} \gamma_i W_i(t) \quad (4.39)$$

Dans la pratique, les modèles non stationnaires *GEV*(μ_t, σ_t, ξ_t) les plus utilisés sont ceux qui traitent (El Adlouni et al., [42]) :

1. Une non stationnarité au niveau de la moyenne avec une variance constante (modèle homoscédastique).
2. Une non stationnarité au niveau de la moyenne et de la variance (modèle hétéroscédastique).

Il faut bien noter que la non stationnarité au niveau de la variance peut être éliminée à l'aide d'une simple transformation (transformation en logarithme par exemple). Nous pouvons citer à titre d'exemple, les modèles GEV non stationnaires suivants :

1. $GEV_0(\mu, \sigma, \xi)$: Les modèles dont les paramètres sont constants. $\mu_t = \mu$, $\sigma_t = \sigma$ et $\xi_t = \xi$. Dans ce cas $n_\mu = n_\sigma = n_\xi = 1$ et $U_1 = V_1 = W_1 = 1$. Ce modèle correspond tout simplement au GEV stationnaire.
2. $GEV_1(\mu_t = \beta_1 + \beta_2 Y_t, \sigma, \xi)$: Les modèles caractérisés par l'existence d'une tendance et d'une homoscédasticité. $n_\mu = 2, U(t) = (U_1(t) = 1, U_2(t) = Y_t), n_\sigma = n_\xi = 1$ et $V_1 = W_1 = 1$.
3. $GEV_2(\mu_t = \beta_1 + \beta_2 Y_t + \beta_3 Y_t^2, \sigma, \xi)$: Les modèles où le paramètre de position est une fonction quadratique d'une covariable temporelle Y_t . Les deux autres paramètres sont constants. Dans ce cas, $n_\mu = 3, U(t) = (U_1(t), U_2(t) = Y_t, U_3(t) = Y_t^2), n_\sigma = n_\xi = 1$ et $V_1 = W_1 = 1$.

Avant d'ajuster un modèle GEV non stationnaire, il faut tout d'abord vérifier la non stationnarité. Nous notons que la structure temporelle (Brunet-Moret (1979, [15])), peut être dégagée par le biais de plusieurs techniques, notamment celles qui se basent sur l'analyse des corrélogrammes des données traitées.

Chapitre 2

Approche des dépassements de seuil : loi GPD

1 Modélisation des excès

Alternativement à l'approche des maxima par bloc, le modèle de renouvellement pour des v.a (i.i.d), a donné naissance à un autre volet de la théorie des valeurs extrêmes, illustré formellement par l'approche des dépassements de seuil (Peaks Over Threshold (POT)). Contrairement à l'approche des maxima par bloc, la méthode POT consiste à utiliser toutes les observations, appelées excès, qui dépassent un certain seuil suffisamment élevé. L'objectif est d'analyser leur comportement asymptotique.

L'approche POT a été introduite pour remédier aux défauts de l'approche classique de la TVE. Les critiques sont essentiellement dues au fait que la méthode d'estimation de la loi GEV se base sur les “ Block component-wise ”, ce qui implique systématiquement une perte d'information. En plus, certains blocs peuvent contenir plusieurs valeurs extrêmes issues de la loi initiale, alors que d'autres peuvent ne pas en contenir (Mc Neil et Frey (2002, [98]), Katz (2002, [76])).

Historiquement, la méthode POT introduite par Pickands (1975, [107]) et reprise par de Haan et Rootzen (1993, [35]), a été initialement appliquée pour le traitement des données hydrologiques, notamment le volume et la durée des déficits en eau. Quant aux volets théoriques de la méthode, ils ont été abondamment développés par divers auteurs tels que Todorovic et Zelenhasic (1970, [142]), Todorovic et Rousselle (1971, [141]), Smith (1987, [131]), Davison et Smith (1990, [33]), Reiss et Thomas (2001, [115]).

Le point de départ de l'approche POT consiste à choisir un seuil $u \in \mathbb{R}$ et de

considérer les variables Y_j définies, à partir des variables initiales X_1, \dots, X_n , par :

$$\begin{cases} N_u = \text{card}\{i : i = 1, \dots, n / X_i > u\} \\ Y_j = X_i - u > 0, \quad \text{pour } 1 \leq j \leq N_u \end{cases} \quad (1.1)$$

Où : X_1, \dots, X_n un échantillon de *v.a* (i.i.d), N_u est le nombre des dépassements du seuil u et Y_1, \dots, Y_{N_u} les excès correspondants.

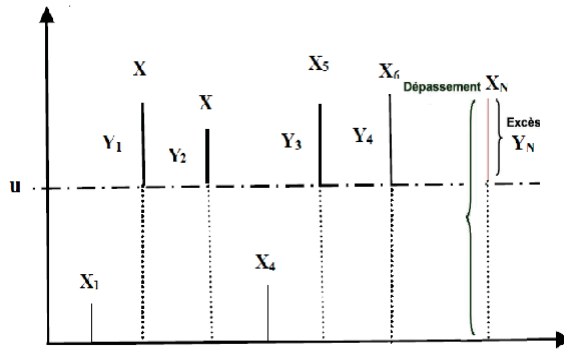


FIGURE 2.1 – Représentation des excès Y issus des dépassements X au-delà d'un seuil u .,(Raggad, 2009 [114])

Le nombre de dépassements d'un seuil u suit une loi de Poisson de paramètre τ , notée $P(\tau)$. Ce résultat est justifié sur la base de l'approximation de la loi binomiale par la loi de Poisson. En effet, si on désigne par $\{X_i > u_n\}$ l'événement "Dépassement d'un seuil u_n " et sa fonction indicatrice $I_{\{X_i > u_n\}}$ définie par :

$$I_{\{X_i > u_n\}} = \begin{cases} 1, & \text{si } X_i > u_n \\ 0, & \text{sinon.} \end{cases} \quad (1.2)$$

alors le nombre de dépassements de u_n , égal à $N_{u_n} = \sum_{i=1}^n I_{\{X_i > u_n\}}$, suit une loi binomiale $\mathfrak{Bin}(n, 1 - F(u_n))$ qui peut être approchée par une loi de Poisson lorsque $n \rightarrow +\infty$, $1 - F(u_n) \rightarrow 0$ et $n(1 - F(u_n)) \rightarrow \tau$.

Une fois les excès identifiés, l'approche POT consiste, en partant de la loi F de X , à déterminer une loi conditionnelle F_u , pour les variables aléatoires Y qui dépassent le seuil u .

La loi conditionnelle F_u est définie par :

$$F_u(y) = \mathbb{P}(X - u \leq y | X > u) = \begin{cases} \frac{F(u+y) - F(u)}{1 - F(u)}, & \text{si } y \geq 0; \\ 0, & \text{si } y < 0. \end{cases} \quad (1.3)$$

La loi asymptotique associée à $F_u(y)$ est donnée par le théorème suivant :

Théorème 1.1 (*Pickands, Balkema et De Haan*) ([107]). *Si F (la loi de X) appartient à l'un des trois domaines d'attraction : \mathcal{D} (Fréchet), \mathcal{D} (Gumbel) ou \mathcal{D} (Weibull), alors il existe une fonction $\sigma(u)$ positive, tel que :*

$$\lim_{u \rightarrow \omega(F)} \sup_{0 < y < \omega(F) - u} |F_u(y) - H_{\xi, \sigma(u)}(y)| = 0 \quad (1.4)$$

où $\omega(F) = \sup\{x \in \mathbb{R} : F(x) < 1\}$ est le point terminal de F et $H_{\xi, \sigma(u)}(y)$ est la fonction de répartition de la loi de Pareto Généralisée (GPD, Generalized Pareto Distribution), définie par :

$$H_{\xi, \sigma(u)}(y) = \begin{cases} 1 - (1 + \frac{\xi y}{\sigma(u)})^{-\frac{1}{\xi}}, & \text{si } \xi \neq 0; \\ 1 - \exp(-\frac{y}{\sigma(u)}), & \xi = 0. \end{cases} \quad (1.5)$$

Pour la démonstration de ce théorème, voir Pickands (1975, [107]).

Nous pouvons estimer la période de retour par la relation suivante (Rosbjerg (1987, [123])) :

$$T_u = \frac{1}{\tau p} = \frac{1}{\tau(1 - H_{\xi, \sigma(u)}(y))} \quad (1.6)$$

Ainsi, si on s'intéresse particulièrement à un niveau de dépassement, une fois chaque N années, nous pouvons utiliser l'équation (1.6) pour déduire que ce niveau de dépassement est donné par :

$$Z_N = \begin{cases} \mu + \frac{\sigma}{\xi} [(N n_y \zeta_u)^\xi - 1], & \text{si } \xi \neq 0; \\ \mu + \sigma \log(N n_y \zeta_u), & \xi = 0. \end{cases} \quad (1.7)$$

où $\zeta_u = \mathbb{P}\{X > u\}$ et n_y est le nombre des dépassements.

Il faut bien signaler que les deux approches ; blocs des maxima et dépassements d'un seuil, sont équivalentes. Ainsi, pour un seuil adéquat, la loi des excès peut être approchée par une loi GPD d'indice extrême ξ , identique à celui de la loi GEV. En effet :

$$\mathbb{P}\{X > x | X > u\} = [1 + \xi(\frac{x - \mu}{\sigma})]^{-\frac{1}{\xi}} \mathbb{P}\{X > x\} = \zeta_u [1 + \xi(\frac{x - \mu}{\sigma})]^{-\frac{1}{\xi}} \quad (1.8)$$

La fonction de répartition ($H_{\xi, \sigma(u)}$) met en exergue trois paramètres : u , σ et ξ . Le paramètre ξ , appelé indice des valeurs extrêmes ou indice de queue, renseigne sur le type de la loi asymptotique (Fréchet pour $\xi > 0$, Gumbel pour $\xi = 0$ et Weibull pour $\xi < 0$). En outre, plus ξ est élevé en valeur absolue, plus le poids des extrêmes est important. Nous parlons dans ce cas d'une loi à queue épaisse.

L'identification du seuil est la phase la plus déterminante dans l'implémentation de l'approche POT, vu que la qualité du modèle en dépend. C'est à partir de la définition d'un bon seuil u que nous pouvons garantir la convergence des excès vers une GPD, estimer les autres paramètres et évaluer le temps de retour.

D'une manière similaire à la détermination des blocs dans la première approche de la TVE, le choix du seuil doit établir un compromis entre biais et variance. Concrètement, le seuil doit être suffisamment grand pour pouvoir utiliser les résultats asymptotiques, mais pas trop élevé afin d'obtenir des estimations précises. Par contre, le choix d'un seuil faible risque de déclarer abusivement des observations extrêmes, introduire un biais dans l'estimation et par conséquent, mal approximer la loi asymptotique. Dans ce sens, plusieurs méthodes de détection du seuil ont été proposées. Certaines ont le défaut d'être subjectives, et donc elles ne doivent être utilisées que pour proposer un intervalle de seuils candidats. D'autres par contre, ont l'avantage d'être objectives.

2 Méthodes pour la détection du seuil

2.1 Méthodes graphiques

2.1.1 Mean Residual Life Plot

La moyenne des excès peut être utilisée pour guider le choix du seuil adéquat (optimal) u^* . En fait, si au dessus d'un seuil u^* , la GPD est une bonne approximation de la loi des excès ($Y = X - u^*$) avec $\xi < 1$, alors :

$$\mathbb{E}[X - u^* | X > u^*] \approx \frac{\sigma_{u^*}}{1 - \xi} \quad (2.9)$$

L'équation (2.9) définit la fonction moyenne des excès notée $e(u^*)$.

Etant donné que l'approximation par une GPD reste valable pour des seuils $u > u^*$, alors :

$$E[X - u | X > u] \approx \frac{\sigma_u}{1 - \xi} \quad (2.10)$$

avec $\sigma_u = \sigma_{u^*} + \xi(u - u^*)$. Ainsi, pour tout $u > u^*$:

$$E[X - u | X > u] \approx \frac{\sigma_{u^*} + \xi(u - u^*)}{1 - \sigma} \quad (2.11)$$

La fonction $e(\cdot)$ est une fonction linéaire de u . Par conséquent, l'identification du seuil u^* consiste à la recherche de la linéarité sur le graphe défini par $(u, e(u))$ et appelé : "mean residual life plot".

En pratique, la fonction $e(\cdot)$ est approchée sur la base de l'approximation empirique $\hat{e}_n(\cdot)$:

$$\hat{e}_n(u) = \frac{\sum_{i=1}^n (X_i - u)}{\sum_{i=1}^n (I_{X_i > u})} \quad (2.12)$$

$$\bar{x}_u = \frac{1}{n_u} \sum x_i (I_{X_i > u}) \quad (u, \bar{x}_u - u)$$



2.1.2 Dispersion Index Plot

u

$$Var[X] = \mathbb{E}[X] \quad (2.13)$$

$$\left\{ \left(u, \frac{\hat{s}_u^2}{\hat{\lambda}_u} \right) : u \leq x_{max} \right\}$$

$$\begin{matrix} \hat{s}_u^2 & \hat{\lambda}_u & & & \\ & u^* & & & \\ & & Var[X] & & \mathbb{E}[X] \end{matrix}$$

2.2 Méthodes numériques

méthodes utilisées pour le choix du seuil optimal, se base sur la minimisation de la variance et du biais, liés aux estimateurs utilisés. Nous allons expliciter quelques unes dans la partie pratique de la présente thèse.

3 Estimation du modèle GPD

3.1 Méthode des moments

la Méthode consiste à égaliser les moments théoriques et les moments empiriques de façon à obtenir :

$$\begin{cases} \mathbb{E}(Y) = \frac{\sigma_u}{1-\xi} = \bar{Y} \\ \mathbb{V}(Y) = \frac{\sigma_u^2}{(1-\xi)^2(1-2\xi)} = S_Y^{*2} \end{cases} \quad (3.14)$$

Où \bar{Y} et S_Y^{*2} représentent respectivement, la moyenne et la variance empirique des excès.

3.2 Méthode du maximum de vraisemblance

La fonction log-vraisemblance est donnée par :

$$\begin{cases} \log(\sigma_u, \xi) = -k \log(\sigma_u) - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^k \log\left(1 + \frac{\xi Y_i}{\sigma_u}\right), & \xi \neq 0; \\ \log(\sigma_u, 0) = -k \log(\sigma_u) - \frac{1}{\sigma_u} \sum_{i=1}^k Y_i, & \xi = 0. \end{cases} \quad (3.15)$$

La résolution numérique du système (3.15) conduit aux estimateurs $\widehat{\sigma}_u^{mv}$ et $\widehat{\xi}_u^{mv}$.

3.3 Méthode des moments de probabilité pondérés

La définition des moments de probabilité pondérés pour une GPD, dans le cas $p = 1$, $r = 0$, $s = 0$ ou $s = 1$, est donnée par :

$$M_{1,0,s} = \frac{\sigma_u}{(s+1)(s+1-\xi)}, \quad \xi < 1 \quad (3.16)$$

Avec :

$$\widehat{M}_{1,0,s} = \frac{1}{k} \sum_{i=1}^k \left(1 - \frac{i}{k+1}\right) Y_{i,k} \quad (3.17)$$

La résolution de l'équation définissant $M_{1,0,s}$ en fonction de σ_u et ξ , conduit pour $s = 0$ et $s = 1$ aux estimateurs *PWM* suivants :

$$\widehat{\xi}_u^{pwm} = 2 - \frac{\widehat{M}_{1,0,0}}{\widehat{M}_{1,0,0} - 2\widehat{M}_{1,0,1}} \quad (3.18)$$

$$\widehat{\sigma}_u^{pwm} = 2 \frac{\widehat{M}_{1,0,0} \widehat{M}_{1,0,1}}{\widehat{M}_{1,0,0} - 2\widehat{M}_{1,0,1}} \quad (3.19)$$

4 Méthodes non paramétriques pour l'estimation de ξ

Une large catégorie d'estimateurs a été élaborée spécifiquement pour le paramètre ξ . Les estimateurs qui font partie de cette catégorie, tel que l'estimateur de Pickands (1975, [107]), Hill (1975, [60]) et Dekkers-Einmahl-De Haan (1989, [38]), sont qualifiés d'estimateurs non paramétriques. Une bonne estimation de ξ peut être obtenue en identifiant une zone de stabilité sur le graphique qui représente différentes valeurs des estimateurs non paramétriques, en fonction du rang k correspondant à une statistique d'ordre $X_{(k)}$ (voir formules des estimateurs).

4.1 L'estimateur de *Pickands*

L'estimateur de Pickands (1975, [107]) combine 4 statistiques d'ordre. Il est calculé pour un ensemble de rang k . Sa formule est donnée par :

$$\widehat{\xi}_{k,n}^P = \frac{1}{\ln 2} \ln \left(\frac{X_{(n-k+1,n)} - X_{(n-2k+1,n)}}{X_{(n-2k+1,n)} - X_{(n-4k+1,n)}} \right) \quad (4.20)$$

Où X_1, \dots, X_n un échantillon de *v.a* (i.i.d) et n est la taille de l'échantillon observé. L'estimateur de Pickands est un estimateur convergent. Plus précisément :

$$\sqrt{k}(\widehat{\xi}_{k,n}^P - \xi) \rightarrow \mathcal{N}(0, \sigma^2(\xi)) \quad (4.21)$$

Où :

$$\sigma^2(\xi) = \frac{\xi^2(2^{2\xi+1} + 1)}{(2(2^\xi - 1) \ln 2)^2} \quad (4.22)$$

4.2 L'estimateur de *Hill*

L'estimateur de Hill (1975, [60]) est l'un des estimateurs les plus répandus dans les applications de la TVE. Introduit par Hill, il est défini de la façon suivante (pour tout $\xi > 0$) :

$$\widehat{\xi}_{k,n}^H = \frac{1}{k} \sum_{i=n-k+1}^n \ln X_{(i,n)} - \ln X_{(n-k+1,n)} \quad (4.23)$$

Le théorème suivant présente les propriétés de l'estimateur de Hill :

Théorème 4.1 ([60]) Soit X_1, \dots, X_n un échantillon de loi F tel que $\bar{F}(x) = x^{-1/\xi}L(x)$, $x > 0$ et L une fonction à variations régulières d'indice 0, alors :

- Si X_1, \dots, X_n un échantillon d'observations i.i.d, $k \rightarrow \infty$ et $k/n \rightarrow 0$ pour $n \rightarrow \infty$, alors l'estimateur de Hill $\hat{\xi}^H$ converge en probabilité vers ξ .
- Si $k/n \rightarrow 0$, $k/\ln(\ln(n)) \rightarrow \infty$ et les X_1, \dots, X_n sont i.i.d, alors l'estimateur de Hill converge presque sûrement vers ξ .
- Si les X_1, \dots, X_n sont i.i.d et si \bar{F} a des variations régulières, alors :

$$\sqrt{k}(\hat{\xi}_{k,n}^H - \xi) \rightarrow \mathcal{N}(0, \xi^2) \quad (4.24)$$

Pour la preuve de ce théorème, voir Hill (1975, [60]).

4.3 L'estimateur des moments de Dekkers – Einmahl – De Haan

L'estimateur de Dekkers-Einmahl-De Haan présente une extension de l'estimateur de Hill pour $\xi \in \mathbb{R}$. Il est défini comme suit :

$$\hat{\xi}_{k,n}^D = 1 + H_n^{(1)} + \frac{1}{2} \left(\frac{(H_n^{(1)})^2}{H_n^{(2)}} - 1 \right)^{-1} \quad (4.25)$$

Où :

$$H_n^{(1)} = \frac{1}{k} \sum_{i=n-k+1}^n \ln X_{(i,n)} - \ln X_{(n-k,n)} \quad (4.26)$$

et :

$$H_n^{(2)} = \frac{1}{k} \sum_{i=n-k+1}^n (\ln X_{(i,n)} - \ln X_{(n-k,n)})^2 \quad (4.27)$$

Sous certaines conditions sur k , l'estimateur de Dekkers-Einmahl-De Haan converge asymptotiquement vers la loi normale.

5 GPD dans le cas des séries non stationnaires

La majorité des chroniques sont marquées par la présence d'un effet saisonnier ou d'une corrélation temporelle. Pour ce type de séries, l'utilisation d'une GPD classique n'est pas valable, vu que les hypothèses d'indépendance et d'homogénéité ne sont pas respectées.

Deux approches sont fréquemment utilisées pour prendre en compte la non stationnarité ou la saisonnalité dans une GPD (Bernard-Michel et al. (2008, [8])). La première consiste à exprimer les paramètres de forme et d'échelle

d'une GPD sous forme d'une fonction de temps (linéaire ou polynomiale pour exprimer une tendance, et sinusoïdale pour illustrer un cycle). La deuxième approche quant à elle, consiste à découper la période d'observation en sous périodes homogènes et ajuster par la suite sur chacune d'elle, une GPD appropriée.

5.1 GPD non stationnaire : paramétrisation de la non stationnarité

La paramétrisation de la non stationnarité signifie que les excès Y_t issus d'une série X_t , suivent une $GPD(\sigma(t), \xi(t))$ dont le paramètre d'échelle σ et de forme ξ dépendent du temps. Dans ce cas, u est noté $u(t)$ pour désigner le seuil à l'instant t .

La principale difficulté de cette approche est liée au choix des fonctions $\sigma(t)$ et $\xi(t)$. La forme linéaire reste la plus simple à utiliser :

$$\begin{cases} \sigma(t) = \exp(\alpha_0 + \alpha_1 t) \\ \xi(t) = \beta_0 + \beta_1 t \end{cases} \quad (5.28)$$

Par ailleurs, la présence d'un cycle correspondant à une période T , peut être prise en considération en ajustant des fonctions de type :

$$\begin{cases} \sigma(t) = \exp(\alpha_0 + \alpha_1 t + \alpha_2(1 - \cos(\frac{2\pi t}{T}))) \\ \xi(t) = \beta_0 + \beta_1 t + \beta_2(1 - \cos(\frac{2\pi t}{T})) \end{cases} \quad (5.29)$$

Une fois $\sigma(t)$ et $\xi(t)$ sont identifiées, les paramètres de forme et d'échelle d'une GPD non stationnaire sont estimés. Cependant, le nombre de paramètres à estimer croît avec la complexité des fonctions choisies. En outre, en augmentant le nombre de paramètres à estimer, nous risquons d'augmenter fortement les incertitudes. Ainsi, le choix des formes de dépendance et d'un seuil en fonction du temps n'est pas une tâche aisée (Coles, (2001, [23]), Khaliq et al. (2006, [77]), Parisi et al. (2000, [104])).

5.2 GPD non stationnaire : définition d'un modèle de mélange

Une deuxième approche pour traiter les séries non stationnaires, consiste à découper la série en blocs sur lesquels l'hypothèse de stationnarité est acceptable (Smith, (1989, [130]), Morton et al. (1997, [100])). Plus concrètement, l'idée consiste à découper l'année en S saisons jugées homogènes. Pour

chaque saison $S_i \in \mathcal{S}$, nous définissons un seuil u_{S_i} au dessus duquel, les excès $Y_j = (X_j - u_{S_i})\mathbf{1}_{X_j \in S_i, X_j > u_{S_i}}$ $GPD(\sigma_{u_{S_i}}, \xi_{u_{S_i}})$

$$h(x) = \sum_{S_i=S_1}^{S_i=S_S} \mathbb{P}(X = x|S_i) \times \mathbb{P}(X \in S_i) \tag{5.30}$$

$$p(X = x|S_i) \quad X$$

x

$$\mathbb{P}(X > x) = \sum_{S_i=S_1}^{S_i=S_S} \mathbb{P}(X > x|X \in S_i) \times \mathbb{P}(X \in S_i) \tag{5.31}$$

$$\mathbb{P}(X > x) = \sum_{S_i=S_1}^{S_i=S_S} \mathbb{P}(X > x|X > u_i, X_i \in S_i) \times \mathbb{P}(X > u_i|X \in S_i) \times \mathbb{P}(X \in S_i) \tag{5.32}$$

$$\widehat{\mathbb{P}}(X > x) = \sum_{S_i=S_1}^{S_i=S_S} \left(1 + \xi_i \frac{x - u_i}{\sigma_i}\right)^{\frac{-1}{\sigma_i}} \times \frac{n_i}{N_i} \times \frac{N_i}{n} \tag{5.33}$$

$$\widehat{\mathbb{P}}(X > x) = \sum_{S_i=S_1}^{S_i=S_S} \left(1 + \xi_i \frac{x - u_i}{\sigma_i}\right)^{\frac{-1}{\sigma_i}} \times \frac{n_i}{n} \tag{5.34}$$

$$n_i \quad N_i$$

$S_i \quad n \quad N$

S

$$K = (N - 1) \frac{\sum_{i=1}^S N_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^S \sum_{j=1}^{N_i} (r_{ij} - \bar{r})} \tag{5.35}$$

$$K = (N - 1) \frac{12}{N(N + 1)} \sum_{i=1}^S N_i (\bar{r}_i - \bar{r})^2 \tag{5.36}$$

Où :

$$\begin{cases} \bar{r}_i = \frac{\sum_{j=1}^{N_j} r_{ij}}{N_i} \\ r = \frac{n+1}{2} \end{cases} \quad (5.37)$$

et :

N_i désigne le nombre d'observations de la saison S_i , n représente le nombre total d'observation sur l'année et r_{ij} désigne le rang de l'observation j de la saison i .

La statistique de Kruskal-Wallis compare le rang moyen des observations de chaque saison au rang moyen sur l'ensemble de la période étudiée (Bernard-Michel et al. (2008, [8])). La valeur de la statistique est plus grande lorsque les saisons sont bien démarquées. Dans le cas contraire, c'est-à-dire quand les saisons se ressemblent, la valeur sera faible. L'idée est donc de maximiser la statistique de Kruskal-Wallis afin de retenir le découpage qui sépare au mieux les saisons.

Après avoir découpé l'année en S saisons, le test de Kruskal-Wallis est effectué pour déterminer si les observations d'une même saison sont *i.i.d.* Si le test sur une saison S_i conduit à une p -value faible, nous pouvons confirmer que les observations de la saison S_i ne sont pas homogènes et par conséquent, le découpage effectué est à rejeter.

Chapitre 3

Adéquation des modèles des valeurs extrêmes

1 Introduction

Quelque soit l'approche adoptée pour la modélisation des valeurs extrêmes, l'analyse de la validité du modèle retenu est jugé nécessaire. Il s'agit de s'assurer que les modèles sélectionnés décrivent bien les données traitées et produisent des bonnes estimations, afin de les employer par la suite dans un processus de décision, notamment pour produire des prévisions et calculer les périodes et les niveaux de retour. Ainsi, des outils de diagnostic et des tests usuels peuvent être utilisés pour juger la pertinence des modèles. Certains outils sont plus spécifiques aux modèles des valeurs extrêmes, d'autres sont communs pour un éventail plus général de modèles.

2 Probabilité et Quantile Plot

Les graphiques Probabilité-Probabilité (P-P plot) ou Quantile-Quantile (Q-Q plot), peuvent être utilisés pour tester l'adéquation des modèles des valeurs extrêmes. Ces outils graphiques simples permettent de comparer la fonction de répartition empirique (pour un P-P plot) ou les quantils (pour un Q-Q plot) d'un échantillon de données à ceux d'un échantillon issu d'une loi théorique (GPD ou GEV par exemple). Si la loi observée est la même que la loi théorique, les points sur ces graphiques doivent être confondus avec la première bissectrice du plan.

3 Tests usuels d'adéquation

Les tests d'adéquation consistent à évaluer les deux hypothèses suivantes :

$$“H_0 : F = F_0” \text{ Versus } “H_1 : F \neq F_0”$$

Où F est la fonction de répartition de l'échantillon (inconnue), et F_0 est une fonction de répartition quelconque (de référence). Il s'agit par la suite, d'identifier une statistique qui soit une mesure raisonnable de l'écart entre F et F_0 , puis de calculer sa loi sous H_0 . Le tableau suivant présente quelques tests d'adéquation, leurs statistiques et formules de calcul.

$$z_j := \widehat{F}_n^{-1}(x_{(j)}), j = 1, 2, \dots, n.$$

KS	$KS := \sqrt{n} \sup_x F_n(x) - F(x) $ $KS = \sqrt{n} \max \left\{ \sup_j \left\{ \frac{j}{n} - z_j \right\}, \sup_j \left\{ z_j - \frac{j-1}{n} \right\} \right\}$
V	$V := \sqrt{n} (\sup_x \{F_n(x) - F(x)\} + \sup_x \{F(x) - F_n(x)\})$ $V = \sqrt{n} \left(\sup_j \left\{ \frac{j}{n} - z_j \right\} + \sup_j \left\{ z_j - \frac{j-1}{n} \right\} \right)$
AD	$AD := \sqrt{n} \sup_x \left \frac{F_n(x) - F(x)}{\sqrt{F(x)(1-F(x))}} \right $ $AD = \sqrt{n} \max \left\{ \sup_j \left\{ \frac{\frac{j}{n} - z_j}{\sqrt{z_j(1-z_j)}} \right\}, \sup_j \left\{ \frac{z_j - \frac{j-1}{n}}{\sqrt{z_j(1-z_j)}} \right\} \right\}$
AD _{up}	$AD_{up} := \sqrt{n} \sup_x \left \frac{F_n(x) - F(x)}{1-F(x)} \right $ $AD_{up} = \sqrt{n} \max \left\{ \sup_j \left\{ \frac{\frac{j}{n} - z_j}{1-z_j} \right\}, \sup_j \left\{ \frac{z_j - \frac{j-1}{n}}{1-z_j} \right\} \right\}$
AD ²	$AD^2 := n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1-F(x))} dF(x)$ $AD^2 = -n + \frac{1}{n} \sum_{j=1}^n (1-2j) \log z_j - \frac{1}{n} \sum_{j=1}^n (1+2(n-j)) \log(1-z_j)$
W ²	$W^2 := n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x)$ $W^2 = \frac{n}{3} + \frac{1}{n} \sum_{j=1}^n (1-2j) z_j + \sum_{j=1}^n z_j^2$
AD _{up} ²	$AD_{up}^2 := n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{(1-F(x))^2} dF(x)$ $AD_{up}^2 = \frac{1}{n} \sum_{j=1}^n (1+2(n-j)) \frac{1}{(1-z_j)} + 2 \sum_{j=1}^n \log(1-z_j)$

FIGURE 3.1 – Quelques tests d'adéquation.

4 Test de rapport de vraisemblance

La vraisemblance (notée dans ce qui suit par \mathcal{L}) est fréquemment utilisée dans le but de quantifier l'adéquation d'une loi de probabilité à un échantillon :

plus grande est la vraisemblance, meilleure est l'adéquation. La particularité des tests de rapport de vraisemblance réside dans le fait que les lois testées doivent appartenir à la même famille. Elles ne diffèrent que par la valeur d'un paramètre θ (vectoriel ou scalaire). H_0 désigne l'ensemble des lois défini par les valeurs du paramètre θ , jugées adéquates à l'échantillon étudié.

Dans le cas où H_0 est vraie, le rapport :

$$\lambda = \max_{H_0} L(x; \theta) / \max_{H_0 \sqcup H_1} L(x; \theta) \quad (4.1)$$

est proche de 1, où $\max_{H_0} L(x; \theta)$ et $\max_{H_0 \sqcup H_1} L(x; \theta)$ désignent les plus grandes valeurs de la vraisemblance donnée par H_0 et $H_0 \sqcup H_1$ respectivement. Dans le cas où H_0 n'est pas vraie, le rapport λ a une valeur proche de 0.

Le test de rapport de vraisemblance est un test statistique utilisé pour comparer la qualité d'ajustement de deux modèles concurrents, dont l'un est un cas particulier de l'autre. Ce test est basé sur un ratio de probabilité utilisé pour calculer une valeur p , afin de pouvoir la comparer avec une valeur critique et décider de rejeter ou non, le modèle sous l'hypothèse nulle en faveur du modèle alternatif. La loi de la statistique du test sous l'hypothèse nulle, peut être approchée en utilisant le théorème de Wilks (Wilks S.S. (1938, [150]) et Chernoff (1954, [20])).

Théorème 4.1 *Si $\Theta_0 \in \Theta$, $\dim \Theta = k$ et $\dim \Theta_0 = I$, et*

$$LR_0 = \frac{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)}{\sup_{\theta \in \Theta} \mathcal{L}(\theta)} \quad (4.2)$$

alors, sous H_0 : " $\theta \in \Theta_0$ ", quand $n \rightarrow \infty$:

$$-2 \log LR_0 \xrightarrow{\mathcal{L}} \chi^2(k - I) \quad (4.3)$$

La région critique du niveau asymptotique α est donnée par :

$$\{-2 \log LR_0 \geq r\} \quad (4.4)$$

Où $\mathbb{P}[\chi^2(k - I) \geq r] = \alpha$

En d'autres termes, le théorème affirme que la statistique du test de rapport de vraisemblance suit une loi de chi carré, avec degré de liberté égal au nombre de paramètres supplémentaires dans le modèle.

5 Exponential Tail “ET” test

Les tests d'adéquation classiques sont influencés par les valeurs les plus probables de la variable traitée. Ils sont conçus pour mesurer l'adéquation aux données de la partie centrale d'une loi, par conséquent, ils ont le défaut de ne pas permettre de juger la qualité de l'extrapolation en queue de la loi étudiée (Garrido, (2002, [53])). Or, cet aspect est le plus important dans le traitement des événements extrêmes. Pour cette raison, de nouveaux tests ont été introduits. Ils sont basés sur la minimisation de l'erreur d'estimation des quantiles extrêmes. L'un de ces tests est l'"Exponential Tail" test, noté “ET” (Garrido, (2002, [53])). Ce test se limite malheureusement aux lois du domaine d'attraction de Gumbel. Pour un quantile extrême fixé d'avance, le “ET” test compare l'estimation donnée par un modèle paramétrique classique et celle induite par l'approche POT. Il faut bien signaler que ce test juge la pertinence du modèle, exclusivement en queue de loi. Dans le cas où le modèle recherché doit s'adapter à l'ensemble des quantiles, un test classique peut suffire.

Soit (X_1, \dots, X_n) un échantillon de taille n , issu de variables aléatoires (*i.i.d*) de fonction de répartition F inconnue.

Soit p_n un nombre positif inférieur à $1/n$. Les quantiles d'ordre $1 - p_n$ issus de F peuvent être estimés d'une façon paramétrique, en utilisant le modèle à tester. Ainsi :

$$\hat{q}_{param,n} = F_{\hat{\theta}_n}^{-1}(1 - p_n) \quad (5.5)$$

Où $\hat{\theta}_n$ est un estimateur des paramètres relatifs à la fonction de répartition F_θ du modèle à tester.

D'une autre manière, les quantiles d'ordre $1 - p_n$ peuvent être estimés d'une façon non paramétrique, en utilisant l'estimateur “ET” (Breiman et al. (1979, [14])), défini par :

$$q_{ET,n} = u_n + \sigma(u_n) \ln\left(\frac{m_n}{np_n}\right) \quad (5.6)$$

Où $\sigma(u_n)$ est une fonction positive et m_n est le nombre d'excès définis au-dessus du seuil u_n . L'idée du “ET” test est de construire un intervalle de confiance basé sur $\hat{q}_{ET,n}$, puis de vérifier si $\hat{q}_{param,n}$ appartient à l'intervalle construit.

Pour un niveau de signification α , l'intervalle théorique asymptotique pour $\hat{q}_{ET,n}$ se définit par :

$$I_{ET,n} = \left[\hat{q}_{ET,n} \pm \sigma_n \frac{\ln(m_n/np_n)}{\sqrt{m_n}} z_{1-\alpha/2} \right] \quad (5.7)$$

L'intervalle de confiance pour le quantile q_{1-p_n} s'écrit :

$$IC_{re,n} = [\hat{q}_{ET,n} + d_n \pm \hat{\sigma}_n \frac{\ln(m_n/np_n)}{\sqrt{m_n}} z_{1-\alpha/2}] \quad (5.8)$$

Où $z_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi normale centrée et réduite.

Chapitre 4

Copules extrêmes

Les relations de dépendance jouent un rôle très important dans de nombreux domaines des mathématiques. Elles constituent l'un des sujets largement étudiés en probabilité et en statistique.

Par ailleurs, pour mesurer la dépendance on a souvent tendance à utiliser le coefficient de corrélation linéaire de Pearson, valable lorsque la relation de dépendance est linéaire et si l'univers considéré est gaussien. De part cette restriction, d'autres limites se posent quant à l'utilisation du coefficient linéaire de Pearson. Nous notons particulièrement qu'il n'est pas approprié pour les lois à queues lourdes, où les variances peuvent être infinies. En plus, il n'est pas invariant par transformation strictement croissante; la corrélation entre X et Y n'est pas la même par exemple entre $\log(X)$ et $\log(Y)$. Par ailleurs, le concept même de la corrélation ne traduit tout simplement qu'une mesure scalaire de la dépendance. Par conséquent, le coefficient ne peut pas renseigner sur la structure de cette dépendance.

En finance comme dans plusieurs domaines, le cas gaussien est rarement vérifié. Pour y remédier, les utilisateurs ont souvent recours à d'autres indicateurs de dépendance, qui se basent sur les discordances et les concordances observées sur un échantillon donné. Nous notons à titre d'exemple, les coefficients de corrélation non linéaires et non paramétriques, comme le tau de Kendall ou le rho de Spearman. Ce sont des bons indicateurs globaux de la dépendance entre variables aléatoires. En outre, ils sont compris entre -1 et $+1$ comme le coefficient de corrélation linéaire, dont la valeur de $+1$ par exemple, traduit une concordance parfaite entre les variables étudiées.

Cependant, mesurer la dépendance à l'aide de quelques indicateurs statistiques est une chose, la modéliser par une fonction de dépendance est une autre. Spécifiquement, dans le contexte des valeurs extrêmes, la modélisation de la

structure de dépendance est préconisée, nous ne pouvons pas nous en passer, par exemple pour traiter le cas multivarié, souvent plus délicat.

C'est ainsi qu'une nouvelle théorie a été introduite. Cette théorie appelée théorie des copules, se présente comme un outil innovant et essentiel à l'appréhension de nombreux domaines d'application, surtout en finance quantitative, telle que la mesure du risque multiple de crédit, l'évaluation des produits de crédits structurés, la réplique de la performance des hedge funds (fonds de couverture), la mesure du risque multiple de marché et la gestion des portefeuilles à l'aide des simulations Monte Carlo.

La majorité des propriétés basiques des copules peut être attribuée aux travaux d'Hoëffding (1940-1941) qui étudia dans le temps les propriétés d'invariance des mesures de dépendance. D'autres propriétés ont été redécouvertes plus tard, plus précisément vers le milieu des années (1970) par Kimeldorf-Sampson (1975), Galambos (1978) et Deheuvels (1978) qui donnèrent aux copules le nom de fonction de dépendance.

Dans la pratique, les champs d'utilisation des copules peuvent être regroupés en deux grandes parties. Tout d'abord, Sklar et Schweizer (1959) se sont intéressés à la théorie probabiliste des espaces métriques. D'ailleurs de 1958 à 1976, la plupart des résultats fondamentaux sur les copules ont été obtenus dans le cadre de l'étude des espaces métriques. Un deuxième champ d'utilisation pourrait porter le nom de mesure de dépendance. Dans ce contexte, Schweizer et Wolf (1981) ont pour la première fois présenté les correspondances entre copule et mesure de dépendance.

Dans la littérature, plusieurs familles de copules ont été proposées, notamment les copules extrêmes qui feront l'objet de ce chapitre. L'intérêt pour ce type particulier de copules se justifie par leur liaison étroite avec la théorie des valeurs extrêmes.

Quant aux applications des copules, nous notons une diversification particulière en finance. Cependant, la théorie qui a été introduite en 1959, n'a pas été appliquée en finance que depuis 1999. Son introduction peut être attribuée à Embrecht, McNeil et Strauman dont les travaux clarifient les concepts essentiels de corrélations au sens large et au sens strict. L'utilisation des copules en finance est préconisée soit pour mesurer la dépendance entre les actifs, soit comme outil de simulations pour évaluer les prix de montages financiers, et dont les valeurs intrinsèques reposent sur les comportements conjoints de plusieurs actifs. Les copules sont utilisées aussi pour comprendre les relations entre les lois conjointes des actifs qui composent un portefeuille ou une politique de

placement.

Les applications en finance peuvent être regroupées en quatre catégories : la gestion des risques (Value-at-risk, risque de défaut, risque de crédit et risque opérationnel), l'évaluation des options, la gestion des portefeuilles et l'étude de l'évolution des co-mouvements des marchés. Nous notons à titre d'exemple les travaux de Cherbuni et Luciano (2000,[19]) qui appliquaient les copules archimédiennes, dans le but d'évaluer les queues de probabilités et la VaR, les recherches de Rosemberg et Schuermann (2006, [125]) qui utilisaient la modélisation par copule pour la prise en compte simultanée des trois types de risques définis par Bâle II, à savoir les risques de marché, de crédit et le risque opérationnel, et enfin, l'article de Li (2000,[86]) qui a étudié le problème de la corrélation du défaut, où il a introduit une variable aléatoire appelée temps avant défaut pour noter le temps de survie de chaque instrument financier.

Il faut bien signaler que l'utilisation des copules n'est pas restrictive à la finance. En science de gestion, la première discipline à avoir adopté les copules était la théorie de l'assurance. L'étude de Tibiletti (2006, [140]) évoque l'utilisation des copules pour modéliser la variation de la demande d'assurance. Yi et Bier (1998, [151]) appliquaient les copules dans le cadre de l'analyse des pré-cureseurs d'accidents. Ils exploraient l'utilisation des copules pour construire les lois de probabilité de faillite sous différents régimes.

En océano-météorologie, l'occurrence simultanée des événements extrêmes peut causer des graves dommages environnementaux ou structuraux. Pour cette raison, il est important de disposer des outils pour construire des modèles qui prennent en compte cette dépendance extrême. Dans ce sens, les copules peuvent bien contribuer à modéliser la dépendance entre les différents paramètres utilisés (hauteur significative et période, houle, vitesse moyenne du vent, etc.). Les copules, dans ce contexte particulier, ont l'avantage de produire une dépendance asymétrique et éventuellement extrême (Marshall et Olkin (1988, [96]), Tawn (1988,1990,[134] et [135]), Joe (1997,[73]), Khoudraji (1998,[78])). Plus récemment, des chercheurs (Pereira et al. (2012,[105])) ont étudié l'apport des copules pour la navigation GPS en présence de multitrajets et en prenant en compte les dépendances spatiales.

1 Présentation mathématique des copules

1.1 Théorème de Sklar et propriétés des copules

Le théorème de Sklar (voir preuve dans Sklar (1959, [129])) est le théorème fondamental dans la théorie des copules. Il permet la liaison entre la densité jointe d'un vecteur aléatoire et ses densités marginales :

Théorème 1.1 (*Théorème de Sklar*) ([129])

Soit \mathbf{H} une loi de probabilité avec des marginales $\mathbf{F}_1, \dots, \mathbf{F}_n$. Il existe une copule \mathbf{C} tel que pour tout $x = (x_1, \dots, x_n) \in \mathbb{R}^n$

$$\mathbf{H}(x_1, \dots, x_n) = \mathbf{C}(\mathbf{F}_1(x_1), \dots, \mathbf{F}_n(x_n)) \quad (1.1)$$

Si les $\mathbf{F}_1, \dots, \mathbf{F}_n$ sont toutes continues, alors \mathbf{C} est unique.

Par le théorème de Sklar, nous pouvons séparer, pour des lois continues, la structure de dépendance de la densité jointe et des densités marginales. Cette structure de dépendance est représentée par la copule.

Un corollaire direct du théorème de Sklar s'énonce comme suit :

Théorème 1.2 ([129]) Soit $\mathbf{H}, \mathbf{C}, \mathbf{F}_1, \dots, \mathbf{F}_n$ les fonctions définies dans le théorème précédent et soient $\mathbf{F}_1^{(-1)}, \dots, \mathbf{F}_n^{(-1)}$ les inverses généralisées de $\mathbf{F}_1, \dots, \mathbf{F}_n$ respectivement. Alors pour tout $\mathbf{u} \in \mathbf{I}^n$ (intervalle de \mathbb{R}^n)

$$\mathbf{C}(u_1, \dots, u_n) = \mathbf{H}(\mathbf{F}_1^{(-1)}(u_1), \dots, \mathbf{F}_n^{(-1)}(u_n)) \quad (1.2)$$

La dérivabilité des copules est justifiée par le théorème suivant :

Théorème 1.3 ([129]) Soit \mathbf{C} une copule. Pour $v \in \mathbf{I}$, la dérivée partielle $\frac{\partial \mathbf{C}}{\partial u}$ existe pour presque tout u , et quand cette dérivée existe, on a

$$0 \leq \frac{\partial \mathbf{C}}{\partial u}(u, v) \leq 1 \quad (1.3)$$

De même, pour tout $u \in \mathbf{I}$, la dérivée partielle $\frac{\partial \mathbf{C}}{\partial v}$ existe pour presque tout v , et quand cette dérivée existe, on a

$$0 \leq \frac{\partial \mathbf{C}}{\partial v}(u, v) \leq 1 \quad (1.4)$$

De plus, les fonctions $u \mapsto \frac{\partial \mathbf{C}}{\partial v}(u, v)$ et $v \mapsto \frac{\partial \mathbf{C}}{\partial u}(u, v)$ sont définies et croissantes presque partout sur \mathbf{I} .

Pour la démonstration des théorèmes et du corollaire, voir Sklar (1959, [129]).

1.2 Exemples de copules

1.2.1 La loi bivariée de Gumbel

Soit \mathbf{H}_θ

$$\mathbf{H}_\theta(x, y) = \begin{cases} 1 - e^{(-x)} - e^{(-y)} + e^{-(x+y+\theta xy)}, & x \geq 0, y \geq 0 \\ 0, & \text{sinon} \end{cases} \quad (1.5)$$

$$\theta \quad [0, 1]$$

$$\mathbf{F}^{(-1)}(u) = -\ln(1-u), \mathbf{G}^{(-1)}(v) = -\ln(1-v) \quad (1.6)$$

$u \ v \in \mathbf{I}$

$$\mathbf{C}_\theta(u, v) = u + v - 1 + (1-u)(1-v)e^{-\theta \ln(1-u) \ln(1-v)} \quad (1.7)$$

1.2.2 La loi bivariée gaussienne

ρ

$$\mathbf{C}_\rho(u, v) = N_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) \quad (1.8)$$

$$\mathbf{C}_\rho(u, v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left(-\frac{1}{2(1-\rho^2)}(s^2 - 2\rho st + t^2)\right) ds dt \quad (1.9)$$

\mathbf{C}_ρ

1.2.3 La famille de Gumbel

$$\Phi(z) = \exp[-z^{1/m}] \quad (1.10)$$

$$\Phi^{-1}(z) = (-\ln z)^m \quad (1.11)$$

$$[-\ln \bar{\mathbf{H}}(x_1, x_2)]^m = [-\ln \bar{\mathbf{H}}_1(x_1)]^m + [-\ln \bar{\mathbf{H}}_2(x_2)]^m \quad (1.12)$$

Soit encore :

$$\bar{\mathbf{H}}(x_1, x_2) = \exp\{-[(-\ln \bar{\mathbf{H}}_1(x_1))^m + (-\ln \bar{\mathbf{H}}_2(x_2))^m]^{1/m}\} \quad (1.13)$$

Où $\bar{\mathbf{H}}$ est la fonction de survie. Ainsi, la copule associée à cette loi est :

$$C(u, v) = \exp\{-[(-\ln u)^m + (-\ln v)^m]^{1/m}\} \quad (1.14)$$

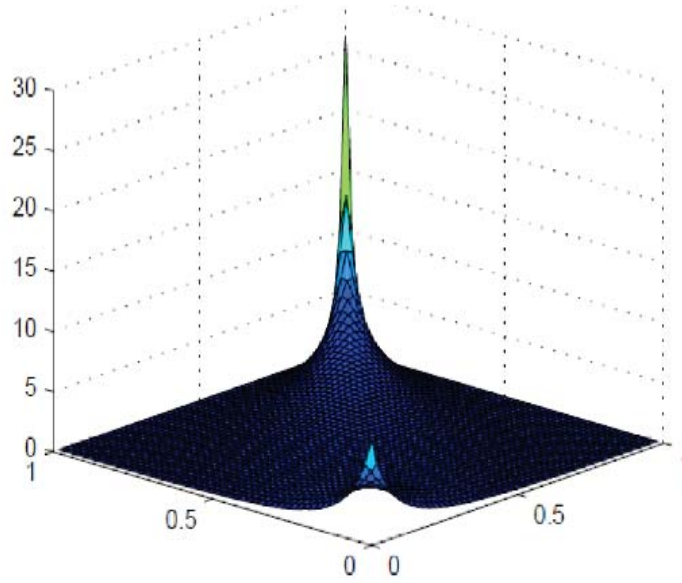


FIGURE 4.1 – Densité de la copule de Gumbel $\theta = 1.5$

1.2.4 La famille de Clayton, Cook et Johnson. G

Cette famille a pour marginales les lois Gamma dont la transformée de Laplace est donnée par :

$$\Phi_\alpha(z) = (1 + z)^{-\alpha} \quad \forall \quad \alpha > 0 \quad (1.15)$$

D'où :

$$\Phi_\alpha^{-1}(z) = z^{-1/\alpha} - 1 \quad (1.16)$$

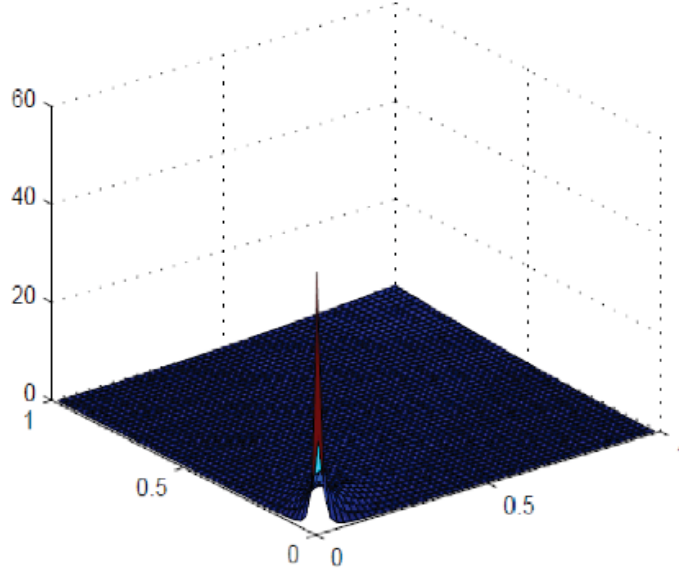
Ainsi :

$$\bar{\mathbf{H}}^{(\alpha)}(x_1, x_2) = \{\bar{\mathbf{H}}_1^{-1/\alpha}(x_1) + \bar{\mathbf{H}}_2^{-1/\alpha}(x_2) - 1\}^{-\alpha} \quad (1.17)$$

$$\bar{\mathbf{H}}^{(\alpha)}(x_1, x_2) = \frac{\bar{\mathbf{H}}_1(x_1)\bar{\mathbf{H}}_2(x_2)}{\{1 - [1 - \bar{\mathbf{H}}_1^{1/\alpha}(x_1)][1 - \bar{\mathbf{H}}_2^{1/\alpha}(x_2)]\}^\alpha} \quad (1.18)$$

La copule qui correspond à cette loi est donnée par :

$$\mathbf{C}(u, v) = \frac{uv}{[1 - (1 - u^{1/\alpha})(1 - v^{1/\alpha})]^\alpha} \quad (1.19)$$

FIGURE 4.2 – Densité de la copule de Clayton $\theta = 2.582$

1.2.5 La famille de Frank

Si on choisit :

$$\Phi(z) = \frac{\ln(1 - \bar{\alpha}e^{-z})}{\ln \alpha}, 0 < \alpha < 1, \bar{\alpha} = 1 - \alpha \quad (1.20)$$

On a alors :

$$\Phi^{-1}(z) = -\frac{\ln(1 - \alpha^z)}{\bar{\alpha}} \quad (1.21)$$

Dans ce cas :

$$\mathbf{H}(x_1, x_2) = \frac{\ln[1 + (\alpha^{\mathbf{H}_1(x_1)-1})(\alpha^{\mathbf{H}_2(x_2)-1})]}{\ln \alpha}, 0 < \alpha < 1 \quad (1.22)$$

On en déduit la copule de Frank :

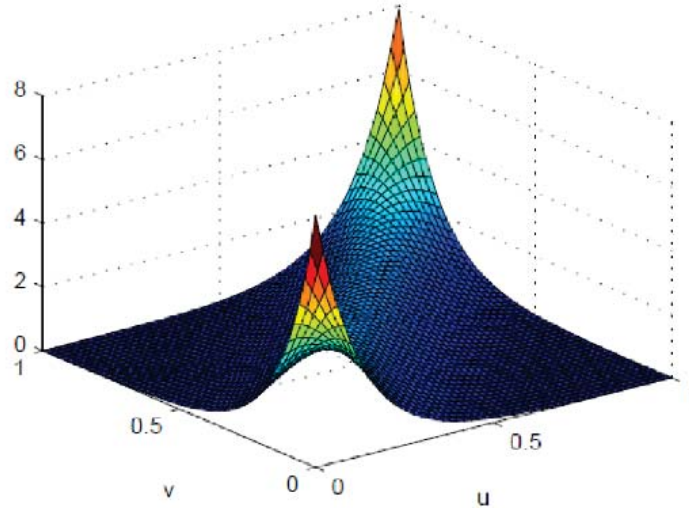
$$C(u, v) = \frac{\ln[1 + (\alpha^u - 1)(\alpha^v - 1)]}{\ln \alpha} \quad (1.23)$$

1.2.6 Les copules archimédiennes

Le théorème suivant permet de définir les copules archimédiennes :

Théorème 1.4 ([45]). Soit Φ une fonction continue, strictement décroissante de $[0, 1]$ dans $[0, \infty]$ et tel que $\Phi(1) = 0$ et soit $\Phi^{[-1]}$ le pseudo-inverse de Φ . La fonction $\mathbf{C} : [0, 1] \times [0, 1] \rightarrow [0, 1]$ définie par :

$$C(u, v) = \Phi^{[-1]}(\Phi(u) + \Phi(v)) \quad (1.24)$$

FIGURE 4.3 – Densité de la copule de Frank $\theta = 7.901$

est une copule si et seulement si Φ est convexe. La fonction Φ est le générateur de la copule.

Pour la démonstration, voir Embrechts et al. (1999, [45]).

Si on prend par exemple $\Phi(t) = (-\ln t)^\theta$ avec $\theta \geq 1$, on obtient la famille de Gumbel :

$$C_\theta(u, v) = \Phi^{[-1]}(\phi(u) + \phi(v)) = \exp\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\} \quad (1.25)$$

1.3 Formes de dépendance

Les copules ont l'avantage de décrire plusieurs formes de dépendance. Nous définissons dans ce qui suit quelques unes :

1.3.1 La dépendance positive (PQD)

Soit (X_1, X_2) un vecteur aléatoire de loi \mathbf{F} et de marginales \mathbf{F}_1 et \mathbf{F}_2 . On dit que \mathbf{F} est **PQD** si :

$$Pr(X_1 \leq a_1, X_2 \leq a_2) \geq Pr(X_1 \leq a_1)Pr(X_2 \leq a_2) \quad \forall a_1, a_2 \in \mathbb{R} \quad (1.26)$$

Ou encore :

$$\mathbf{C}(\mathbf{F}_1(a_1), \mathbf{F}_2(a_2)) \geq \mathbf{F}_1(a_1)\mathbf{F}_2(a_2) \quad \forall a_1, a_2 \in \mathbb{R} \quad (1.27)$$

1.3.2 La dépendance positive stochastique croissante (DPSC)

Soit (X_1, X_2) un vecteur aléatoire de loi \mathbf{F} et de marginales \mathbf{F}_1 et \mathbf{F}_2 . On dit que X_2 est stochastiquement croissante en X_1 si :

$$Pr(X_2 > x_2 | X_1 = x_1) = 1 - \mathbf{F}_{2|1}(x_2 | x_1) \quad (1.28)$$

est croissante en x_1 pour tout x_2 . De même, en inversant le rôle des variables on peut définir la croissance stochastique de X_1 en X_2 . La croissance stochastique signifie que la valeur de l'une des variables a plus de chance de croître lorsque l'autre croît.

1.3.3 La dépendance de queue

Le concept de dépendance de queue renseigne sur "l'intensité" de dépendance au niveau des queues de la loi étudiée. C'est un outil très pertinent pour l'étude de la dépendance des valeurs extrêmes.

Soient X et Y deux variables aléatoires continues de lois \mathbf{F}_1 et \mathbf{F}_2 respectivement.

Le coefficient de dépendance de queue supérieure de X et Y est défini par (pourvu que la limite $\lambda_U \in [0, 1]$ existe) :

$$\lambda_U = \lim_{u \rightarrow 1^-} Pr[Y > F_2^{-1}(u) | X > F_1^{-1}(u)] \quad (1.29)$$

Si $\lambda_U \in (0, 1]$, X et Y sont dites asymptotiquement dépendantes au niveau de la queue supérieure de la loi étudiée. Si $\lambda_U = 0$, X et Y sont dites asymptotiquement indépendantes au niveau de la queue supérieure de la loi étudiée.

L'indice de queue inférieure est défini par :

$$\lambda_L = \lim_{u \rightarrow 0^+} \frac{\mathbf{C}(u, u)}{u} \quad (1.30)$$

Si $\lambda_L > 0$, cela signifie qu'il y a une probabilité non nulle que l'une des variables dépasse u sachant que l'autre est déjà supérieure à u .

Généralement, nous n'appliquons pas directement cette formule de queue. Nous utilisons le fait que pour une paire de variables aléatoires (U, V) dont chaque composante suit une loi uniforme $[0, 1]$, on a :

$$\begin{cases} Pr[V \leq v | U = u] = \frac{\partial}{\partial u} \mathbf{C}(u, v) \\ Pr[V > v | U = u] = 1 - \frac{\partial}{\partial u} \mathbf{C}(u, v) \end{cases} \quad (1.31)$$

Ainsi :

$$\lambda_U = \lim_{u \rightarrow 1^-} (Pr[V > u|U = u] + Pr[U > u|V = u]) \quad (1.32)$$

De plus, si la copule est symétrique, c'est à dire si $\mathbf{C}(u, v) = \mathbf{C}(v, u)$, on obtient :

$$\lambda_U = 2 \lim_{u \rightarrow 1^-} Pr[V > u|U = u] \quad (1.33)$$

Pour le cas gaussien par exemple, on a :

$$\lim_{u \rightarrow 1^-} Pr[V > u|U = u] = \lim_{x \rightarrow \infty} Pr[\Phi^{-1}(V) > x|\phi^{-1}(U) = x] = Pr[X > x|Y = x] \quad (1.34)$$

Or la loi de $Y|X = x$ est une normale $N(\rho x, 1 - \rho^2)$. On en déduit que :

$$\lambda_U = 2 \lim_{x \rightarrow \infty} \bar{\Phi}((x - \rho x)\sqrt{1 - \rho^2}) = 2 \lim_{x \rightarrow \infty} \bar{\Phi}(x\sqrt{1 - \rho}/\sqrt{1 + \rho}) \quad (1.35)$$

Donc pour $\rho < 1$, la copule gaussienne n'indique pas qu'il y a une dépendance au niveau de la queue supérieure puisque $\lambda_U = 0$. On obtient un résultat analogue dans le cas d'une copule de Student (Embrechts, McNeil et Straumann (1999,[45]) :

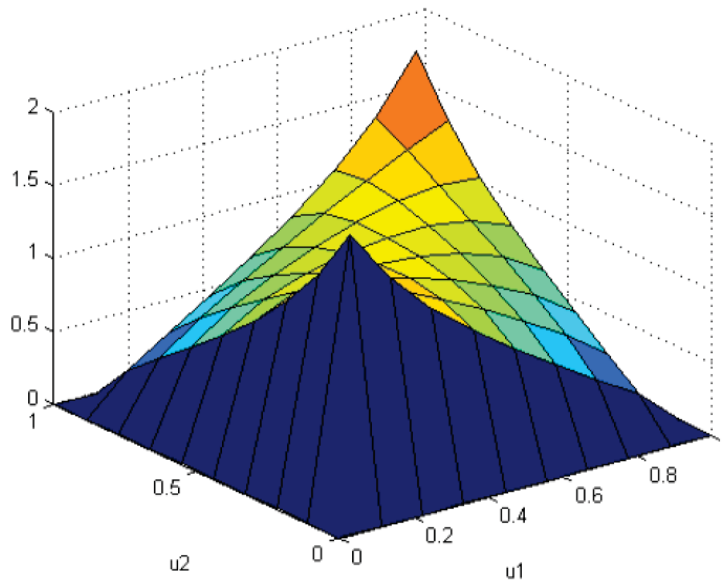
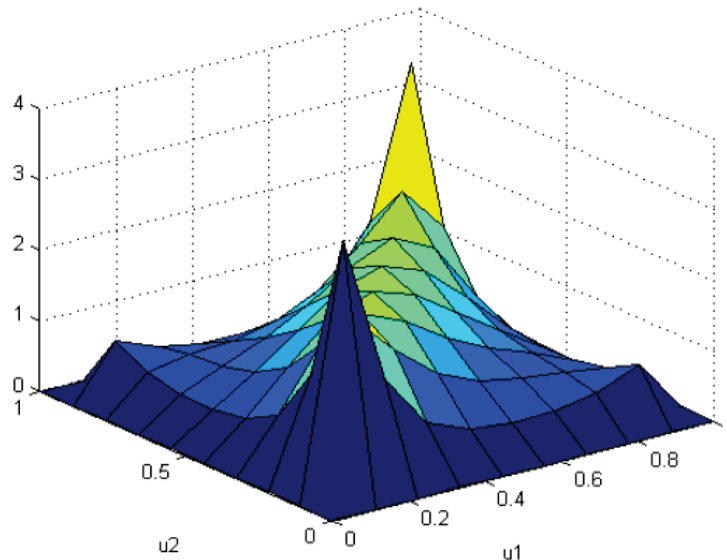


FIGURE 4.4 – Densité de la copule gaussienne avec $\rho = 0,5$.

$$\lambda_U = 2\bar{t}_{\nu+1}(\sqrt{\nu+1} - \rho/\sqrt{1+\rho}) \quad (1.36)$$

Dans ce cas, λ_U est égal à 0 uniquement si $\rho = -1$. On remarque aussi que cette dépendance décroît avec le degré de liberté ν de la copule. Pour $\rho < 1$, on

FIGURE 4.5 – Densité de la copule de Student avec $\rho = 0.5$ et $\nu = 1$.

remarque que les extrêmes ne sont pas corrélés lorsque la structure de dépendance est gaussienne, alors qu'ils le sont lorsque la dépendance est une copule de Student. Ce résultat est très important pour la modélisation financière, il met en exergue le fait que si la normalité est considérée comme une hypothèse valable, un problème se pose donc lors de l'agrégation des mesures de risques.

2 Les copules extrêmes

Considérons un vecteur aléatoire (X, Y) de loi H et de marginales continues H_1 et H_2 . D'après le théorème de Sklar, il existe une unique fonction copule tel que :

$$H(x, y) = C(H_1(x), H_2(y)) \quad (2.37)$$

Considérons une suite de vecteurs aléatoires $(X_1, Y_1), \dots, (X_n, Y_n)$ indépendants et de lois $H(x, y)$. Soit

$$(X_n^*, Y_n^*) = (\max(X_1, \dots, X_n), \max(Y_1, \dots, Y_n)) \quad (2.38)$$

On dit que H appartient au domaine d'attraction d'une loi des valeurs extrêmes H^* avec des lois marginales non dégénérées, s'il existe des suites $a_{1n} > 0$, $a_{2n} > 0$, b_{1n} et b_{2n} tel que :

$$\lim_{n \rightarrow \infty} Pr\left[\frac{X_n^* - b_{1n}}{a_{1n}} \leq x, \frac{Y_n^* - b_{2n}}{a_{2n}} \leq y\right] = \lim_{n \rightarrow \infty} H^n(a_{1n}x + b_{1n}, a_{2n}y + b_{2n}) = H^*(x, y) \quad (2.39)$$

Soit C^* la copule correspondante à H^* et soient H_1^* et H_2^* les marginales de H^* . H appartient au domaine d'attraction de H^* si et seulement si H_1 et H_2 appartient aux domaines d'attraction de H_1^* et H_2^* respectivement et si C satisfait :

$$C^*(x, y) = \lim_{n \rightarrow \infty} C(x^{1/n}, y^{1/n})^n, \forall 0 \leq x, y \leq 1 \quad (2.40)$$

3 Les copules archimax

Cette nouvelle famille de copules a été introduite par Capéraà, Fougères et Genest (2000,[17]). Elle englobe la plupart des familles de copules connues, notamment les copules archimédiennes et les copules extrêmes. Cette nouvelle famille offre plus de flexibilité pour la modélisation.

Nous signalons qu'une fonction bivariée est une copule archimax si et seulement si elle est de la forme :

$$C_{\Phi, A}(x, y) = \Phi^{-1}[(\Phi(x) + \Phi(y))A(\frac{\Phi(x)}{\Phi(x) + \Phi(y)})] \quad (3.41)$$

pour tout $0 \leq x, y \leq 1$, avec :

1. $A : [0, 1] \rightarrow [1/2, 1]$ tel que $\max(t, 1-t) \leq A(t) \leq 1$ pour tout $0 \leq t \leq 1$;
2. $\Phi : (0, 1] \rightarrow [0, \infty)$, est une fonction convexe, décroissante qui vérifie $\phi(1) = 0$. On adoptera la convention $\phi(0) = \lim_{t \rightarrow 0^+} \phi(t)$ et $\phi^{-1}(s) = 0$ pour $s \geq \phi(0)$.

Ainsi, si on remplace x et y par des lois univariées F et G , on obtient une loi bivariée de marginales F et G . Cette nouvelle famille de copules contient toutes les copules archimédiennes et toutes les copules extrêmes. En effet, en choisissant $\Phi(t) = \ln(1/t)$, on obtient :

$$C_{\phi, A}(x, y) = C_A(x, y) = \exp[\ln(xy)A(\frac{\ln(y)}{\ln(xy)})] \quad (3.42)$$

pour tout $0 \leq x, y \leq 1$.

Plus particulièrement, si $A(t) \equiv 1$, on retrouve la forme générale des copules archimédiennes :

$$C_{\phi, A}(x, y) = C_{\phi}(x, y) = \phi^{-1}(\phi(x) + \phi(y)) \quad (3.43)$$

4 Méthodes d'estimation

4.1 Les copules empiriques

Les copules empiriques ont été introduites par Deheuvels (1979,[37]).

Soit $\chi = (x_1^t, \dots, x_N^t)_{t=1}^T$ un échantillon. La copule empirique est donnée par :

$$\widehat{C}\left(\frac{t_1}{T}, \dots, \frac{t_N}{T}\right) = \frac{1}{T} \sum_{t=1}^T \mathbf{1}_{[x_1^t \leq x_1^{(t_1)}, \dots, x_N^t \leq x_N^{(t_N)}]} \quad (4.44)$$

Où $x_n^{(t)}$ représente la statistique d'ordre et $1 \leq t_1, \dots, t_N \leq T$. La fréquence de la copule empirique est :

$$\widehat{c}\left(\frac{t_1}{T}, \dots, \frac{t_N}{T}\right) = \frac{1}{T} \quad (4.45)$$

si $(x_1^{(t_1)}, \dots, x_N^{(t_N)}) \in \chi$.

La relation entre la copule empirique et sa fréquence est donnée par :

$$\widehat{C}\left(\frac{t_1}{T}, \dots, \frac{t_N}{T}\right) = \sum_{i_1=1}^{t_1} \dots \sum_{i_N=1}^{t_N} \widehat{c}\left(\frac{t_1}{T}, \dots, \frac{t_N}{T}\right) \quad (4.46)$$

Les copules empiriques peuvent être utilisées pour estimer les mesures de dépendance. Par exemple, une estimation du rho de Speraman dans le cas bivarié est donnée par :

$$\widehat{\rho}_S = \frac{12}{T^2 - 1} \sum_{t_1=1}^T \sum_{t_2=1}^T \left(\widehat{C}\left(\frac{t_1}{T}, \frac{t_2}{T}\right) - \frac{t_1 t_2}{T^2} \right) \quad (4.47)$$

4.2 Estimation par la méthode du maximum de vraisemblance

Soit θ un vecteur de paramètres, de dimension $K \times 1$ et Θ l'espace dans lequel θ prend ses valeurs. La vraisemblance pour l'observation t est notée $L_t(\theta)$. Désignons par $l_t(\theta)$ la log-vraisemblance. Pour T observations, $\widehat{\theta}_{ML}$ est l'estimateur du maximum de vraisemblance si :

$$l(\widehat{\theta}_{ML}) \geq l(\theta) \quad \forall \theta \in \Theta \quad (4.48)$$

Avec $l(\theta) = \sum_{t=1}^T l_t(\theta)$ est la log-vraisemblance correspondante à T observations.

$\widehat{\theta}_{ML}$ est asymptotiquement normal et on a :

$$\sqrt{T}(\widehat{\theta}_{ML} - \theta_0) \rightarrow N(0, \mathcal{I}^{-1}(\theta_0)) \quad (4.49)$$

avec $\mathfrak{J}^{-1}(\theta_0)$ la matrice d'information de Fischer.

Puisque la densité de $\mathbf{F}(x_1, \dots, x_N) = C(F_1(x_1), \dots, F_N(x_N))$ est

$$f(x_1, \dots, x_N) = c(\mathbf{F}_1(x_1), \dots, \mathbf{F}_N(x_N)) \prod_{n=1}^N f_n(x_n) \quad (4.50)$$

où f_n et $c(u_1, \dots, u_n) = \frac{\partial C}{\partial u_1, \dots, \partial u_n}$ sont les densités respectives de \mathbf{F}_n et de la copule, alors on a :

$$l(\theta) = \sum_{t=1}^T \ln c(\mathbf{F}_1(x_1^t), \dots, \mathbf{F}_N(x_N^t)) + \sum_{t=1}^T \sum_{n=1}^N \ln f_n(x_n^t) \quad (4.51)$$

Chapitre 5

Modélisation des extrêmes pour la gestion des risques hydro-météorologiques

1 Gestion des risques hydro-météorologiques

Le changement climatique est la principale menace environnementale actuelle. Si la température du globe continue à croître, les événements exceptionnels telles que les inondations, les tempêtes, les vagues de chaleur, les coulées de boue ou les sécheresses, vont nécessairement se multiplier. Par conséquent, les phénomènes extrêmes de nature dévastateurs, auront un impact négatif sur plusieurs domaines. L'augmentation de la température provoquera une plus grande évaporation qui va faire augmenter la quantité des précipitations. L'approvisionnement alimentaire sera perturbé et donc, de vastes régions vont devenir inaptes à produire la nourriture. L'eau constituera un bien rare qui va générer des conflits et ces faits ne seront pas sans conséquences sur la stabilité des systèmes politiques des régions.

Pour suivre ces évolutions, les cercles académiques s'intéressent à concocter des modèles capables de prévoir de la manière la plus fiable, les potentielles occurrences des événements extrêmes dans les années à venir. A partir des données disponibles, ces modèles offrent la possibilité de décrire et d'effectuer une analyse des phénomènes inhabituels. L'analyse de ces phénomènes permet de déterminer la probabilité des événements plus extrêmes que ceux d'ores et déjà observés, et d'estimer leurs niveaux qui seraient susceptibles d'être atteints dans une période 5, 10, ou même 100 fois plus grande que la période d'observation. Les résultats de cette analyse conditionnent la prise de décision.

Sur un site maritime par exemple, prévoir les hauteurs des vagues extrêmes sur de longues périodes est primordial pour le dimensionnement des ouvrages portuaires. Dans le cas de la construction d'un barrage hydro-électrique qui occasionne des frais gigantesques, la prise en considération du niveau maximal qui sera probablement atteint, est extrêmement importante. Elle permet d'une part d'optimiser les coûts de construction, et d'autre part de prémunir les populations environnantes contre les dommages et les risques. En outre, le calcul des crues extrêmes est nécessaire pour la protection des zones industrielles, les populations et leurs habitations. Il permet la définition de l'échelle d'intensité des événements extrêmes, la sécurité des ouvrages hydrauliques et la validité des plans de prévention des risques.

En fait, la conception des ouvrages hydrauliques et la gestion des ressources hydriques est essentiellement conditionnée par l'estimation des débits des crues. Ainsi, l'hydrologue est toujours amené à modéliser le processus de crue, phénomène complexe et variable dans le temps et dans l'espace. Pour ce faire, les hydrologues font appel à des modèles qui s'inscrivent dans le cadre de l'analyse fréquentielle. Il s'agit de plusieurs approches statistiques de prévision qui consistent à étudier les événements passés afin d'en définir les probabilités d'apparition future. Nous pouvons citer à titre d'exemple, le modèle de dépassement appelé modèle des séries chronologiques de durées partielles. Ce dernier modèle comporte toutefois une difficulté majeure lors de son application ; il s'agit de déterminer préalablement le seuil de troncature, défini comme le débit minimal des crues.

Dans le même contexte, nous présentons dans la deuxième partie de ce chapitre un travail qui a été réalisé et publié dans *Journal of Hydrologic Engineering* (Managing hydrological risks with extreme modeling : application of peaks over threshold model to the loukkos basin watershed, 2014).

L'étude examine les outils objectifs permettant l'identification du seuil et par conséquent, la détermination de la meilleure loi qui ajuste les extrêmes des débits hydrologiques.

2 Application du POT aux données hydrologiques

2.1 Introduction

L'application de l'approche POT à un échantillon des excès conduit à l'ajustement d'une GPD. Récemment, plusieurs approches ont été développées pour estimer les paramètres qui en résultent, notamment le paramètre du seuil μ et de forme ξ . Nous suggérons dans cette partie une mise en œuvre des approches qualifiées de méthodes numériques, et utilisées pour la détection du seuil μ de la GPD. Dans la première étape, nous introduisons l'utilisation du test du rapport de vraisemblance (LRT) pour ajuster la forme de la loi : ($\xi = 0$) ou ($\xi \neq 0$). Dans la deuxième étape, nous comparons trois méthodes de détection du seuil. Il s'agit du test du rapport de vraisemblance (LRT), la méthode basée sur la minimisation de l'erreur quadratique (SEM) et la méthode du seuil multiple (MTM). Les lois qui en résultent seront appliquées pour la gestion des risques hydrologiques. Le cas traité se rapporte aux débits d'une ressource hydrique d'une grande importance au Maroc. Il s'agit du bassin du Loukkos situé dans la région du Nord.

2.2 Données

Le Loukkos constitue un maillon essentiel du Nord-Ouest marocain et de son développement. C'est une région qui dispose de plusieurs atouts considérables, dues essentiellement à sa position stratégique qui l'entoure de grands centres de consommation et la place près de l'Europe, avec 2 ports importants à Laarache et à Tanger. En outre, le milieu physique de la région présente un avantage majeur avec des reliefs diversifiés, des précipitations importantes ainsi qu'une main d'œuvre jeune et abondante.

La zone du Loukkos se divise en deux bassins ; le Loukkos et le Dradère-Soueire. Le premier

recouvre les bassins hydrologiques qui prennent naissance dans la chaîne rifaine : oueds Loukkos, Ouarour et oued Makhazine. Par ailleurs, la région du Loukkos bénéficie d'un climat méditerranéen caractérisé par l'alternance de deux saisons. L'une est humide et fraîche et l'autre, sèche et chaude. En outre, la pluviométrie moyenne annuelle qui s'élève à 700 mm, est plus concentrée pour sa quasi-totalité sur la période allant d'octobre à avril. Ces particularités ont permis un développement du secteur agricole dans la région qui s'est bien



FIGURE 5.1 – Présentation du bassin du Loukkos

accéléral ces dernières années, notamment avec des productions végétale et animale particulièrement élevées, ainsi que de nombreux investissements directs étrangers.

Cependant, les valeurs moyennes ne reflètent en rien le caractère aléatoire et contrasté du climat de la région. Le régime des précipitations se caractérise par des contrastes à la fois saisonniers et annuels, ce qui conduit à un régime hydrologique marqué par la violence et la rapidité des averses, les fortes pentes et l'imperméabilité des terrains ainsi que des inondations de type crues brutales. Les variations climatiques contribuent pour une grande part à ces catastrophes, qui pourront être la conséquence de deux principales composantes. D'une part, l'implantation de toutes sortes de constructions, d'équipements et d'activités dans l'espace alluvial, progressivement façonné par le cours d'eau, aggrave le risque de vulnérabilité et par conséquent, d'inondation. D'autre part, la présence occasionnelle de l'eau dans ce qu'on appelle en hydrologie, par opposition au lit mineur ; le lit majeur, modifie les conditions d'écoulement de l'eau et conduit à de graves catastrophes.

Pour se prémunir contre ces risques, une politique de gestion a été mise en place. Elle consiste à préparer des plans opérationnels qui s'articulent au tour de la prévention et l'aménagement. La prévention débute par la compréhension de l'aléa, son étude et son suivi à l'aide d'un système d'information fiable. Pour ce faire, il est primordial donc de collecter les statistiques nécessaires à l'appréhension de tels phénomènes extrêmes.

Dans le même ordre d'idées, nous proposons dans cette partie d'étudier les débits mensuels du Loukkos fournis par la Direction Régionale de l'Hydrologie et enregistrés, à la station Medouar pendant la période allant de 1969 à 1998. Des données supplémentaires (de 1999 à 2006) vont être utilisées pour la calibration du modèle élaboré.

Afin d'utiliser le théorème de Balkema-de Haan-Pickands, pour modéliser la queue de la loi des données hydrologiques, issues du bassin du Loukkos, nous devons tester la validité des hypothèses d'application, notamment que les observations sont i.i.d. Puisque la loi est inconnue, les tests non paramétriques d'homogénéité et d'indépendance ont été utilisés. Leur application ne confirme pas la validité de la condition i.i.d. Afin de répondre aux exigences théoriques, nous considérons la série des maxima trimestriels (116 observations) qui satisfont bien, selon les mêmes tests, l'homogénéité et l'indépendance.

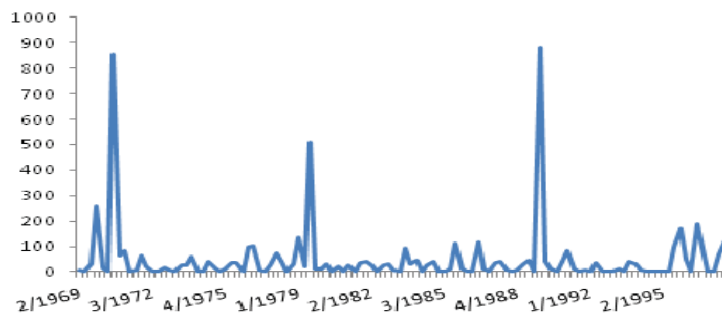


FIGURE 5.2 – Evolution des débits trimestriels du Loukkos pendant la période allant du deuxième trimestre 1969 au premier trimestre 1998

La figure ci-dessus met en évidence la nature volatile des données hydrologiques traitées, justifiant ainsi l'intérêt d'étudier et par la suite de prévoir l'occurrence de leurs valeurs extrêmes. Par ailleurs, l'histogramme de la série montre que la loi des données est étalée vers la droite. En outre, le coefficient d'asymétrie est de l'ordre de 5 et la valeur du coefficient d'aplatissement indique clairement, que les données suivent une loi leptokurtique.

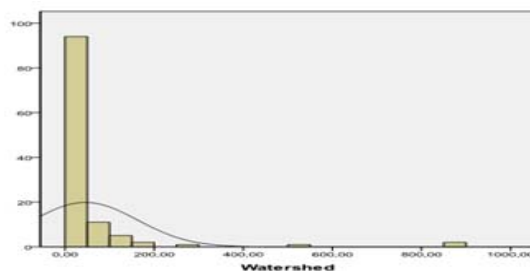


FIGURE 5.3 – Histogramme des débits trimestriels du Loukkos

2.3 Méthodologies

Soit Z_1, \dots, Z_n une suite d'observations i.i.d, issue d'une loi de probabilité inconnue F . La loi des excès $X_i = Z_i - u$ au-dessus du seuil u est donnée par :

$$F_u(x) = \mathbb{P}\{Z - u \leq x \mid Z > u\} = \mathbb{P}\{X \leq x \mid Z > u\} = \frac{F(x + u) - F(u)}{1 - F(u)} \quad (2.1)$$

La loi des excès peut être approximée par une $GPD(u, \sigma_u, \xi)$. Où u est le seuil, ξ est le paramètre de forme et σ_u le paramètre d'échelle. u est considéré comme un **seuil admissible** si la GPD ajuste bien F_u . Un seuil admissible u est dit **optimal**, si u est le plus petit seuil admissible qui maximise la p-value d'un test d'adéquation.

Par ailleurs, les approches qui ont été développées pour identifier le seuil optimal, peuvent être regroupées en deux catégories : la première se base sur des considérations graphiques et présente par conséquent, une grande part de subjectivité. La seconde quant à elle, utilise des outils analytiques. Nous présentons dans ce qui suit les principales méthodes que nous avons adoptées pour traiter les données hydrologiques du Loukkos.

MRL plot et test de rapport de vraisemblance pour ξ

Afin de pouvoir estimer les paramètres de la GPD, il faut s'assurer de l'adéquation de cette loi pour ajuster la loi d'un échantillon donné. Ainsi, il faut d'une part, sélectionner un seuil u au dessus duquel un nombre de données soit conservé pour produire des estimations précises et d'autre part, il faut choisir un seuil suffisamment élevé pour que l'approximation par la GPD soit applicable. Il faut bien rappeler, que si nous choisissons un seuil trop bas, les estimations seront biaisées, par contre, si le seuil est trop élevé, les estimations seront imprécises, puisque leurs écarts types seront très grands.

Pour un choix judicieux du seuil, nous adoptons dans une première étape, une méthode graphique basée sur le "MRL" plot. Ce graphique représente les points suivants :

$$\{(u, e_n(u)), Z_{n:n} < u < Z_{1:n}\} \quad (2.2)$$

Où : $Z_{1:n} = \inf\{Z_i\}$, $Z_{n:n} = \sup\{Z_i\}$ et $e_n(u)$ est défini par :

$$e_n(u) = \frac{\sum_{i=1}^n (Z_i - u)^+}{\sum_{i=1}^n 1_{\{Z_i > u\}}} = \frac{1}{N_u} \sum_{i=1}^n (Z_i - u)^+ \quad (2.3)$$

Où $(Z_i - u)^+ = \sup(Z_i - u, 0)$ et $e_n(u)$ désigne l'estimateur empirique de

la moyenne des excès notée $e(u)$ avec :

$$e(u) = \mathbb{E}[Z - u | Z > u] \quad (2.4)$$

Le "MRL" plot des données issues d'une GPD, possède un comportement linéaire au-dessus d'une certaine valeur de u . Cette constatation est justifiée par le fait que la moyenne des excès pour une GPD est donnée par :

$$e(u) = \frac{\sigma_u + \xi u}{1 - \xi} \quad (2.5)$$

Rappelons que si la loi des excès au-dessus d'un seuil u_1 est une $GPD(u_1, \sigma_{u_1}, \xi)$, alors pour tout seuil $u_2 > u_1$, la GPD reste valable, avec le même paramètre de forme ξ et un paramètre d'échelle σ_{u_2} donné par :

$$\sigma_{u_2} = \sigma_{u_1} - \xi(u_2 - u_1) \quad (2.6)$$

Ainsi, cette relation est vérifiée pour tous les seuils admissibles. Pour un seuil optimal u^* et pour tout $u > u^*$ nous avons :

$$\mathbb{E}(X - u | X > u) = \frac{\sigma_u}{1 + \xi} = \frac{\sigma_{u^*} - \xi(u - u^*)}{1 + \xi} \quad (2.7)$$

Pour une meilleure exploration du "MRL" plot, nous aurons besoin d'estimer ξ . Puisque la loi n'est pas connue d'avance, il s'avère nécessaire d'utiliser des estimateurs non paramétriques tel que l'estimateur des moments ou de Hill.

L'estimateur des moments et le "MRL" plot, permettent donc d'avoir une idée sur la queue de la loi et le signe de ξ . Ils permettent en plus, la détermination d'un intervalle de seuils candidats. Ainsi, pour choisir le seuil optimal, nous aurons besoin d'utiliser, dans une deuxième étape, des méthodes analytiques. Nous utilisons, en particulier, un test de rapport de vraisemblance pour déterminer laquelle des deux hypothèses est juste :

$$H_0 : \xi = 0 \text{ ou } H_1 : \xi \neq 0$$

Ce qui permettra par la suite d'identifier la loi de probabilité à tester :

$$F_u(x) = F_u(x, \sigma_u, \xi) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi}, & \text{si } \xi \neq 0; \\ 1 - \exp(-x), & \text{si } \xi = 0. \end{cases} \quad (2.8)$$

Une fois la forme de la loi qui ajuste au mieux les données est déterminée, nous pouvons passer à l'étape suivante qui consiste à identifier le seuil optimal u^* . Pour ce faire, nous proposons l'utilisation et la comparaison de plusieurs méthodes analytiques, à savoir la méthode SEM, la méthode MTM et la méthode basée sur LRT.

Square Error Method (SEM)

Le choix du seuil pourra être effectué, comme proposé par Beirlant et al. (1996, [4]), sur la base de la minimisation de l'erreur quadratique moyenne (EQM), d'un estimateur de l'indice de queue, notamment l'estimateur de Hill. L'erreur quadratique moyenne est utile dans le sens qu'elle permet de comparer plusieurs estimateurs, en particulier lorsque l'un d'eux est biaisé. Il est donc tout à fait évident que le seuil optimal à retenir, est celui qui minimise l'EQM d'un estimateur basé sur les dépassements (Guillou et Willems (2006, [56])). Quant à l'estimation de l'EQM, Beirlant et al. (1999, [3]) la calculent en utilisant une approximation de la fonction de répartition asymptotique de $\xi_{k,n}$. Danielsson et al. (2001, [28]), Draisma et al. (1999, [32]), Gomes et Oliveira (2001, [55]) suggèrent d'utiliser des méthodes de bootstrap. Pour le choix du seuil correspondant à nos données, nous adaptons le travail de Beirlant en estimant d'une façon empirique, l'EQM. Le seuil optimal à retenir est en fin de compte celui qui minimise l'EQM selon les étapes décrites par l'algorithme suivant :

Soient u_1, \dots, u_n les n seuils équidistants obtenus selon une approche graphique tel que MRL plot. Pour $j = 1, \dots, n$, nous notons $\hat{\sigma}_{u_j}$ et $\hat{\xi}_{u_j}$ les estimateurs des paramètres d'échelle σ_{u_j} et de forme ξ_{u_j} , basés sur l'échantillon des excès au delà du seuil u_j :

Etape 1 Déterminer l'effectif N_{u_j} des excès au delà du seuil u_j .

Etape 2 Simuler ν échantillons indépendants de taille N_{u_j} , issus d'une loi $GPD(u_j, \hat{\sigma}_{u_j}, \hat{\xi}_{u_j})$.

Etape 3 Pour chaque $\alpha \in A = \{0.05, 0.1, 0.15, \dots, 0.95\}$ et pour chaque $i = 1, \dots, \nu$, calculer les quantiles $q_{(\alpha, u_j)}^i$ du i^{me} échantillon simulé. Calculer ensuite $q_{(\alpha, u_j)}^{\text{sim}} = \frac{1}{\nu} \sum_{i=1}^{\nu} q_{(\alpha, u_j)}^i$.

Etape 4 Pour chaque $j = 1, \dots, n$, calculer l'erreur quadratique moyenne défini par :

$$SE_{u_j} = \sum_{\alpha \in A} (q_{(\alpha, u_j)}^{\text{sim}} / q_{(\alpha, u_j)}^{\text{obs}})^2 \quad (2.9)$$

Où $q_{(\alpha, u_j)}^{\text{obs}}$ est le quantile observé, analogue au quantile basé sur les échantillons simulés ; $q_{(\alpha, u_j)}^{\text{sim}}$.

Le seuil optimal u^* retenu sur la base de l'application de l'algorithme SEM est en fin de compte, celui qui correspond à : $SE_{u^*} = \min_j SE_{u_j}$.

Multiple Threshold Method (MTM)

La MTM a été développée par Deidda (2010, [30]) pour déduire les paramètres de la GPD, ajustée aux dépassements des relevés pluviométriques quotidiens dans diverses régions, et qui correspondent par conséquent à différents seuils. Le choix de cette méthode est justifié par ses performances lorsqu'il s'agit de traiter des données sujettes à différentes sources de perturbation, des relevées irrégulières ou dont les caractéristiques changent d'un site à l'autre. Ainsi, la MTM basée sur le concept des paramètres invariants, est beaucoup plus appropriée aux comparaisons et analyses des extrêmes régionaux, où les seuils optimaux dépendent fortement des sites observés. Ce concept de seuil invariant, est obtenu sur la base des équations suivantes :

Supposons que la loi des excès au delà d'un seuil $u \geq 0$ est donnée par :

$$F_u(x) = F_u(x, \sigma_u, \xi) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi}, & \text{si } \xi \neq 0; \\ 1 - \exp(-x), & \text{si } \xi = 0. \end{cases} \quad (2.10)$$

En utilisant l'équation :

$$\sigma_{u_2} = \sigma_{u_1} - \xi(u_2 - u_1) \quad (2.11)$$

reliant les paramètres d'échelle σ_{u_1} et σ_{u_2} , correspondant aux seuils admissibles u_1 et u_2 avec $u_2 > u_1$, nous obtenons :

$$\sigma_0 = \sigma_u - \xi u \quad (2.12)$$

et par conséquent :

$$\forall u \geq 0, \quad F(u) = (1 - \zeta_0) + \zeta_0 F_0(u) \quad (2.13)$$

et :

$$\forall u \geq 0, \quad \zeta_u = \zeta_0 [1 - F_0(u)] \quad (2.14)$$

Ainsi :

$$\forall u \geq 0, \zeta_0 = \begin{cases} \zeta_u (1 + \xi_u \frac{u}{\sigma_0})^{\frac{1}{\xi}} = \zeta_u (1 - \xi_u \frac{u}{\sigma_u})^{-\frac{1}{\xi}}, & \text{si } \xi_u \neq 0; \\ \zeta_u \exp \frac{u}{\sigma_0} = \zeta_u \exp \frac{u}{\sigma_u}, & \text{si } \xi_u = 0. \end{cases} \quad (2.15)$$

Avec $\zeta_0 = P\{X > 0 | X \geq 0\}$ et $\zeta_u = P\{X > u | X \geq 0\}$.

La dernière équation met en exergue la reparamétrisation ζ_0 , qui illustre le concept de seuil invariant (Deidda (2010, [30])).

Les étapes d'application de la MTM sont synthétisées par le biais de l'algorithme suivant :

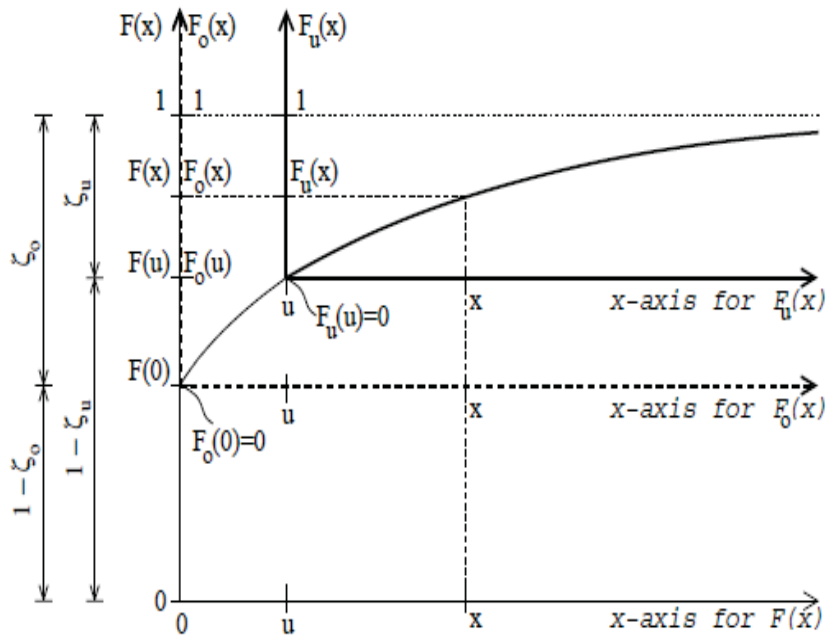


FIGURE 5.4 – Illustration des différentes relations utilisées pour la mise en application de MTM

Etape 1 (L'estimateur ξ^M) : Identifier n seuils candidats : $u_1 < u_2 < \dots < u_n$. L'estimateur ξ^M correspond à la médiane des estimateurs ξ_{u_j} basés sur les échantillons des dépassements au-delà de u_j .

Etape 2 (L'estimateur σ_0^M) : Afin de neutraliser la variabilité de l'estimateur de σ due aux fluctuations induites par la variabilité des ξ , nous estimons dans l'étape 2, le paramètre d'échelle sous condition que $\xi = \xi^M$. Pour chaque échantillon, on le note $\sigma_{u_j}^c$. L'estimateur final du paramètre d'échelle est obtenu, d'une façon semblable à celle de l'étape 1, comme la médiane des estimateurs $\sigma_{u_j}^c$ basés sur les échantillons des dépassements au-delà de u_j .

Etape 3 (L'estimateur ζ_0^M) : Identiquement aux étapes 1 et 2, ζ est estimé sous conditions que $\xi = \xi^M$ et $\sigma = \sigma_0^M$. Pour chaque échantillon, on note l'estimateur de ζ par $\zeta_{u_j}^c$. L'estimateur est obtenu comme la médiane des estimateurs $\zeta_{u_j}^c$ basés sur les échantillons des dépassements au-delà de u_j .

L'estimation finale de ζ permet d'identifier le nombre adéquat de dépassements et par conséquent, le seuil optimal.

La méthode basée sur LRT pour le choix du seuil optimal

Le test de rapport de vraisemblance peut être utilisé pour identifier le seuil optimal. Soient deux seuils admissibles u_i et u_k avec $u_i < u_k$. Soient \mathbb{X}_{u_i} (respectivement \mathbb{X}_{u_k}) le vecteur des dépassements au delà du seuil u_i (respectivement u_k). Soient $\hat{\xi}_{u_i}$ et $\hat{\sigma}_{u_i}$ (respectivement $\hat{\xi}_{u_k}$ et $\hat{\sigma}_{u_k}$) les estimateurs de maximum de vraisemblance des paramètres de forme et d'échelle, basés sur \mathbb{X}_{u_i} (respectivement \mathbb{X}_{u_k}).

La recherche du seuil optimal consiste à parcourir l'intervalle des seuils admissibles en testant :

$$(H_0 : u_i \text{ est le seuil optimal}) \text{ vs } (H_1 : u_i \text{ n'est pas le seuil optimal})$$

La statistique du test correspondante est donnée par :

$$LR(u_i; u_k) = -2 \log \frac{f(\hat{\xi}_{u_i}, \hat{\sigma}_{u_i}, \mathbb{X}_{u_i})}{f(\hat{\xi}_{u_k}, \hat{\sigma}_{u_k}, \mathbb{X}_{u_k})} \quad (2.16)$$

Avec : f est la fonction de vraisemblance.

Le comportement asymptotique de la statistique LR est décrit par Wilks (1938, [150]) et Chernoff (1954,[20]), qui montrent que sous H_0 , LR suit asymptotiquement une loi χ^2 .

La mise en œuvre de la méthode basée sur LRT est effectuée selon l'algorithme suivant :

Etape 1 : Choisir un seuil u_k "suffisamment élevé" tel que u_{k-1} soit un seuil admissible et le nombre des excès soit supérieur à 30. Cette dernière condition assure les caractéristiques asymptotiques du LRT.

Etape 2 : Calculer la p -value de la statistique $LR(u_{k-1}, u_k)$ pour tester ($H_0 : LR(u_{k-1}, u_k)$ suit une χ^2)

Etape 3 : Si l'hypothèse H_0 est rejetée avec un risque α fixé d'avance, substituer k par $k - 1$ et répéter l'étape 2.

Le seuil optimal est en fin de compte, celui qui correspond à la p -value la plus grande, calculée au niveau de l'étape 2 de l'algorithme ci-dessus. Par ailleurs, nous pouvons initialiser l'algorithme par un seuil u_k qui correspond à la médiane des seuils admissibles sélectionnés sur la base du MRL plot. En outre, l'adéquation de la loi χ^2 pourra être effectuée par le biais d'un test d'ajustement tel que le test d'Anderson-Darling.

2.4 Résultats

En se basant sur le MRL plot et la représentation de l'estimateur non paramétrique de Dekkers, en fonction de plusieurs valeurs de K , nous pouvons constater une stabilité des estimateurs pour l'intervalle $[17, 40]$. Cet intervalle constitue donc l'ensemble des seuils candidats pour les données du bassin du Loukkos. En outre, la pente positive du "MRL" plot permet de préjuger que la loi de Férchet pourra ajuster au mieux les excès des données du Loukkos.

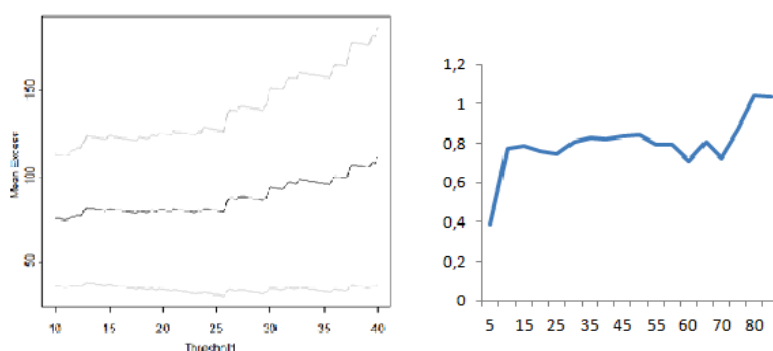


FIGURE 5.5 – Présentation du MRL plot et de l'estimateur de Dekkers en fonction de plusieurs valeurs de K , pour les débits trimestriels du Loukkos

La détermination de la forme de la loi est basée sur le LRT. Ce test appliqué à plusieurs échantillons de taille variant entre 31 et 56 et identifiés selon les seuils candidats, confirme le rejet de l'hypothèse nulle ($H_0 : \xi = 0$). En effet, $LR = 5,31 > 3,84$.

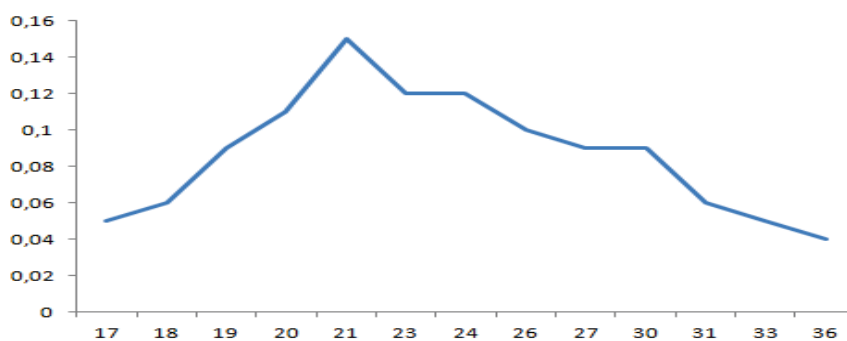


FIGURE 5.6 – Les p-values du test d'Anderson Darling, correspondantes aux différents seuils admissibles

En outre, l'application du LRT pour la détection du seuil, conduit à l'identification de plusieurs seuils admissibles. Cependant, le test d'Anderson Darling,

TABLE 5.1 – Les lois résultantes des différentes méthodes MTM, SEM et LRT

Méthode	Seuil optimal	Loi ajustée	test AD2up
LRT	21	$F_{21}(x, 31.25, 0.65) = 1 - (1 + 0.65 \frac{x-21}{31.25})^{-\frac{1}{0.65}}$	0.15
MTM	24	$F_{24}(x, 27.17, 0.75) = 1 - (1 + 0.75 \frac{x-24}{27.17})^{-\frac{1}{0.75}}$	0.12
SEM	23	$F_{23}(x, 26.38, 0.75) = 1 - (1 + 0.75 \frac{x-23}{26.38})^{-\frac{1}{0.75}}$	0.05

conduit à confirmer que la valeur 21 possède la plus grande p-value parmi tous les seuils admissibles du LRT. Ainsi, selon LRT et le test d'Anderson Darling, nous pouvons conclure que le seuil optimal correspondant à la série des débits trimestriels du Loukkos est le débit de 21 mm.

Les autres méthodes; MTM et SEM, permettent d'ajuster des lois de Fréchet, mais qui diffèrent en terme des paramètres estimés.

L'adéquation de ces lois selon les tests d'ajustement AD2up et le QQ plot, montre que la méthode LRT fournit le meilleur ajustement.

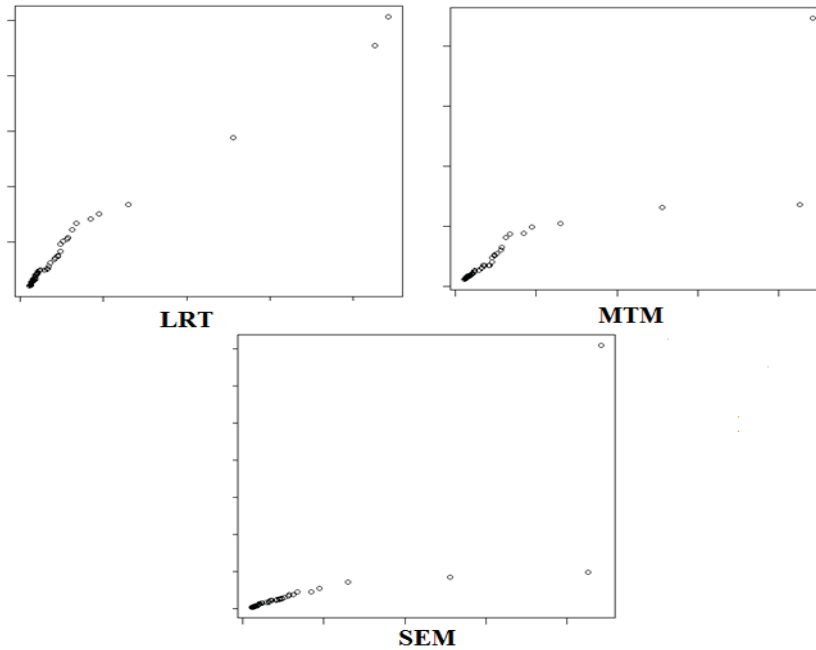


FIGURE 5.7 – Le QQ-plot correspondant aux différentes lois ajustées

En se basant sur ces résultats graphiques et analytiques, nous pouvons ajuster aux excès des débits du Loukkos, la loi de Fréchet suivante :

$$F_{21}(x, 31.25, 0.65) = 1 - (1 + 0.65 \frac{x - 21}{31.25})^{-\frac{1}{0.65}} \quad (2.17)$$

Nous pouvons caractériser donc la loi des débits du Loukkos par sa queue

lourde, ce qui signifie en pratique pour ce versant bassin, la fréquence d'atteindre avec forte chance des seuils élevés. Cette fréquence de dépassement justifie en grande partie, la multiplication des plans d'aménagement de la région du Loukkos qui visent entre autres, la régularisation des eaux pour l'irrigation, la contribution à la protection de la plaine de Loukkos et la ville de Ksar El Kébir contre les inondations, ainsi que la production de l'énergie électrique. En outre, l'agence des bassins hydrauliques du Loukkos a inventorié les sites à risque dans sa zone d'action (plus de 80 sites sont considérés comme des sites à risque). Ces sites ont fait l'objet de classement selon le degré de gravité des inondations qu'ils subissent, ainsi que le degré d'urgence de l'intervention prévue, déterminé à la base de l'ampleur des dégâts constatés ou potentiels.

Pour tester la performance du modèle ajusté, nous allons vérifier dans un premier temps sa stabilité. Dans un deuxième temps, le modèle sera utilisé pour effectuer des prévisions en se basant sur la notion du temps de retour. L'étude de stabilité de notre modèle s'est portée sur les débits maximums trimestriels du bassin du Loukkos, observés sur la période allant de 2^{ème} trimestre 1969 au 4^{ème} trimestre 2006. Une première partie, en particulier la période allant du 2^{ème} trimestre 1969 au 1^{er} trimestre 1998, a été utilisée pour l'ajustement de la loi $F_{21}(x, 31.25, 0.65)$. La deuxième période qui s'étale sur le 2^{ème} trimestre 1998 au 4^{ème} trimestre 2006 sera consacrée pour calibrer le modèle. Dans le cas où la stabilité est vérifiée, le modèle sera utilisé pour faire des prévisions. La figure ci-dessous révèle des niveaux élevés des débits pendant la seconde période retenue. Ceci est dû en grande partie au fait que, au cours des années 2000, la région du Loukkos a connu plusieurs inondations, notamment en 2000 et 2002. Par ailleurs, en appliquant le modèle précédent sur toute la période

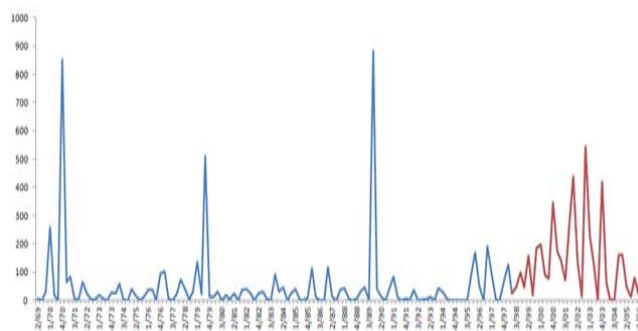


FIGURE 5.8 – Evolution des débits du Loukkos pour la période allant du 2^{ème} trimestre 1969 au 4^{ème} trimestre 2006

TABLE 5.2 – Ré-estimation pour la période initialement retenue

Paramètres	$2^{eme}trim1969$ au $1^{er}trim1998$	$2^{eme}trim1969$ au $4^{eme}trim2006$
Seuil	21	23
Paramètre d'échelle	31.25	45.97
Paramètre de forme	0.65	0.63

d'étude, nous avons obtenu une p-value du test AD2up de l'ordre de 12%. En outre, la méthode *LRT*, appliquée à nouveau à toute la période d'étude, fournit des estimations très proches de celles de la période initialement retenue. Ces résultats prouvent que le modèle reste valable pour les huit années qui suivent la période d'élaboration du modèle (2^{eme} trimestre 1969 au 1^{er} trimestre 1998). Par conséquent, nous pouvons utiliser ce modèle pour calculer le temps de retour d'une valeur extrême. En se basant sur l'analyse des données hydrologiques, nous avons constaté que la région du Loukkos a connu une fréquence croissante des inondations depuis 1996, notamment celles de 2000 et 2002, où les débits maximaux trimestriels ont atteint respectivement 342 mm et 542 mm. Pour ces deux événements, le modèle GPD estime un temps de retour de l'ordre de 23 trimestres et 44 trimestres respectivement. Ces valeurs estimées sont très proches des valeurs 21.42 et 50 trimestres, calculées sur la base des durées marquées par les débits réellement observés. Nous pouvons conclure donc, que la performance de notre modèle retenu est clairement mise en évidence. Ainsi, ce modèle peut servir d'une grande part, à gérer les risques hydrologiques au niveau de la région du Loukkos.

Chapitre 6

Gestion des mouvements extrêmes : Crises financières, crashes boursiers et chocs pétroliers

1 Mouvements extrêmes liés aux marchés boursiers

La bourse joue un rôle primordial dans l'économie contemporaine. Elle se présente comme une composante du marché des capitaux, qui permet la valorisation des actifs financiers qui y sont cotés et par conséquent, le financement de l'économie. Chaque actif financier a un prix qui dépend des anticipations des investissements sur les revenus futurs qu'il peut rapporter. Grâce au marché boursier, les entreprises peuvent trouver une partie des capitaux nécessaires à leur expansion, l'Etat peut financer le déficit de ses comptes et les investisseurs peuvent faire fructifier leur épargne, en devenant associés d'entreprises ou créanciers des collectivités publiques.

Ainsi, le marché boursier est scindé en deux sous marchés qui sont respectivement : le marché primaire et le marché secondaire. Le premier marché concourt à la mutation des structures de production des entreprises, permettant ainsi à celles désireuses de trouver des financements, de rencontrer des détenteurs de capitaux. Tandis qu'au deuxième marché ; le marché secondaire, il offre aux détenteurs des produits financiers la possibilité de s'échanger les titres en fonction de leur stratégie d'investissement. En tant que marché primaire, la bourse

permet de transformer l'épargne des ménages en ressources longues pour le compte des collectivités publiques et privés. En tant que marché secondaire, le marché financier assure la liquidité et la mobilité de l'épargne. Le marché secondaire, où les produits financiers répondent à la loi de l'offre et de la demande, est de loin le plus actif des deux marchés.

Le marché boursier est caractérisé par des fortes fluctuations. Ces mouvements spectaculaires qui exposent tout intervenant à un risque de perte, ont ravivé chez les autorités de régulation et de contrôle, les praticiens et les cercles académiques, l'intérêt à la question de la volatilité des marchés financiers. La multiplication des chocs exceptionnels tels que la crise asiatique en 1997, la crise russe et son impact sur les pays émergents en 1998, la crise du fonds spéculatif LTCM la même année, les événements du 11 septembre 2001 et la crise argentine en 2001, a profondément modifié la perception du risque des investisseurs et les perspectives de croissance. Actuellement, nous traversons une crise financière bouleversante, caractérisée par des hausses et des baisses déconcertantes qui se répercutent sur l'ensemble des marchés financiers.

Une telle situation a suscité l'intérêt de plusieurs académiciens et théoriciens qui tentent d'explorer l'efficience de la théorie des valeurs extrêmes à décrire et à prévoir l'évolution des marchés financiers. Nous pouvons citer, à titre d'exemple, l'article de Longin F. ([87]) intitulé "The Choice of the Distribution for Asset Returns : Extreme Values Can Help?", où l'auteur utilise les résultats concernant la loi des extrêmes pour discriminer entre différentes lois statistiques pour des rentabilités journalières. La théorie des valeurs extrêmes a été utilisée aussi en finance, pour l'étude des séries des taux de change (Holes et de Vries (1991, [61]), Jansen et de Vries (1991, [70])) et pour celles des rendements sur les marchés des actions (Loretan et Philipps (1994, [92]), Longin (1996, [89]), Jondeau et Rockinger (1999, [74])). Une synthèse des résultats théoriques et des aspects économétriques peut être trouvée dans Embrechts, Kluppelberg et Mikosch (1997, [44]).

Dans la même optique, nous présentons un article que nous avons publié dans *International Journal of Statistics and Economics* (Statistical methods to expect extreme values : Application of POT, approach to CAC40 return index, 2012). L'article illustre l'application de la théorie des valeurs extrêmes pour modéliser les rendements de l'indice CAC40 afin de mieux gérer les cours extrêmes. L'analyse des rentabilités semble être devenue un préambule incontournable à toute décision d'investissement. Cette analyse apparaît d'autant plus justifiée que les investisseurs et les managers du risque deviennent parti-

culièrement sensibles aux événements se réalisant sous des conditions extrêmes (Scott et Horvath (1980, [128]), Pratt et Zeckhauser (1987, [109])).

L'intérêt des études dans ce domaine, est de sensibiliser les investisseurs au risque potentiel de leur placement financier. Ils pourront ainsi, en fonction de leurs contraintes et leur aversion au risque, se couvrir ou se prémunir contre une probable réalisation des événements extrêmes.

2 Chocs pétroliers et impacts économiques

Le pétrole est l'un des moteurs fondamentaux de la croissance économique mondiale. Il est utilisé directement ou indirectement dans un grand nombre de processus de transport et de la production industrielle. Par conséquent, les variations de son prix ont un impact marqué sur la conjoncture économique. Plus encore, les retombées s'avèrent d'autant plus importantes car il n'existe pas de bien de substitution à court terme au pétrole, et qui peut répondre rapidement aux mêmes besoins à un prix plus faible.

Un choc pétrolier se définit comme un phénomène de hausse brutale du prix du pétrole, tout en affectant profondément les économies à l'échelle mondiale. Selon les économistes, deux chocs pétroliers ont marqué l'histoire du *XX^e* siècle : le premier en 1973, le second en 1979. Par extension, certains spécialistes, journalistes et hommes politiques, considèrent l'augmentation des prix du pétrole à partir de la seconde moitié des années 2000 et son paroxysme en 2008, comme un troisième choc pétrolier.

Plusieurs facteurs peuvent être évoqués pour expliquer un choc pétrolier. Cependant, ces facteurs qui varient d'une crise à l'autre peuvent être illustrés, dans la majorité des cas, par le déséquilibre entre l'offre et la demande ainsi que par les tensions géopolitiques. Nous pouvons parler ainsi, d'un choc d'offre, qui lui même peut être dû à une crise politique ou un conflit armé dans un pays ou un ensemble de pays producteurs ou de transit, telle que la crise de 1973.

Nous pouvons citer aussi, les baisses volontaires de l'offre des pays producteurs et l'augmentation non anticipée de la demande des pays consommateurs illustrant ce que l'on appelle un choc de demande.

Un choc pétrolier engendre d'abord une phase de ralentissement de la croissance, durant laquelle les ménages consomment moins, ce qui entraîne une diminution de la production des entreprises. Ce ralentissement est source d'une augmentation du chômage et donc par effet de chaîne, d'une baisse encore ac-

crue du pouvoir d'achat. Les périodes de crise pétrolière peuvent engendrer des périodes de récession durant lesquelles le Produit Intérieur Brut (PIB) d'un pays ralentit sa croissance ou même diminue. La diminution de l'activité engendre une déstabilisation économique, liée essentiellement aux transferts monétaires massifs des pays importateurs de pétrole vers les pays exportateurs.

Dans un contexte affecté par les chocs pétroliers et un marché fondamentalement volatil, il est primordial de disposer des outils permettant de mieux quantifier les risques liés aux fluctuations des prix du pétrole brut, et qui sont en réalité difficiles à prévoir, puisqu'ils dépendent des facteurs géopolitiques, des aléas climatiques, des perspectives économiques et de l'évolution des capacités de production. Dans ce sens, la théorie des valeurs extrêmes peut apporter un éclairage pour aider les investisseurs et les gestionnaires des risques, qui sont particulièrement sensibles à l'occurrence des pertes substantielles. En plus, sachant l'interdépendance entre chocs pétroliers et plusieurs variables macroéconomiques, une variété de travaux a été développée pour capter la structure de dépendance en se basant sur la théorie multivariée des valeurs extrêmes.

3 Gestion des mouvements extrêmes

L'économie mondiale connaît de plus en plus divers changements et mutations. Ainsi, nous assistons à une extrême libéralisation des économies émergentes et à une variété de fusions et de regroupements au niveau des économies industrialisées. Il va sans dire qu'un tel processus de libéralisation, dans le contexte d'une mondialisation intensifiée, risque de bouleverser l'économie mondiale et par conséquent, il est plus que jamais dans l'intérêt des gestionnaires des portefeuilles et des investisseurs en général, de gérer convenablement les risques extrêmes.

Le risque est inhérent à la plupart des activités humaines, surtout lorsqu'il s'agit du secteur bancaire. Cependant, jusqu'à la fin des années 80, le risque bancaire prenait essentiellement deux formes : le risque de crédit (ou de contrepartie) qui correspond aux risques encourus par l'établissement financier et qui sont liés, à la défaillance d'une contrepartie ou/et à la dégradation de la qualité du crédit, et le risque de liquidité qui illustre la difficulté potentielle de réaliser rapidement une transaction. Par ailleurs, la libéralisation des mouvements internationaux des capitaux et le développement croissant, des instruments

dérivés sur les marchés financiers durant les années 90, ont été accompagnés par une augmentation considérable de l'activité de trading dans les banques. Ainsi, d'autres formes de risques ont été envisagées. Nous pouvons évoquer à cet égard, le risque de marché et le risque opérationnel. Le premier désigne le risque des pertes sur les positions du bilan et du hors bilan suite aux variations des prix de marché tels que les taux d'intérêt, les prix des actions, les cours de change et les cours de matières premières. Tandis que le deuxième ; risque opérationnel, il se définit comme le risque des pertes dû à une inadéquation ou à une défaillance des procédures, du personnel, des systèmes internes ou à des événements extérieurs.

Dans la pratique, la diversification des risques pose un problème méthodologique important qui se rapporte à la recherche d'une mesure objective et pertinente du risque. Cette tâche semble de moins en moins aisée, surtout pour des domaines où la qualification du risque est plus floue et où les données sont moins nombreuses.

En particulier pour le domaine financier, le risque est traditionnellement associé à l'incertitude, basée sur la variation des prix des actifs et mesurée par la volatilité. Cette mesure est couramment utilisée dans la théorie du portefeuille de Markowitz. Cependant, la mesure du risque financier par la volatilité est basée sur la convergence en loi normale, qui n'est pas valable pour les queues des lois, où se situent les risques les plus importants. D'ailleurs, dans le calcul de la variance, les valeurs extrêmes sont souvent considérées comme aberrantes et par conséquent, elles sont exclues. Pour ladite raison, la réglementation bancaire s'attache à d'autres mesures, plus aptes à contrôler les événements extrêmes qui font peser un réel danger sur les établissements financiers.

Quelle que soit la mesure de risque adoptée, les praticiens constatent que, pour les marchés bancaires et financiers, les niveaux de couverture sont particulièrement élevés et par conséquent, les événements qui correspondent à ces niveaux ont une faible probabilité d'occurrence. Pour cette raison, les mesures de risque sont généralement complétées dans la pratique par des programmes de scénarios de crise (stress testing), qui permettent de déterminer le montant de perte auquel la banque doit faire face lors d'une prochaine crise, si le portefeuille de négociation reste inchangé.

Les autorités réglementaires, les établissements financiers doivent effectuer régulièrement des simulations de crise afin de connaître le montant des pertes potentielles en cas des fluctuations dangereuses et importantes du marché. Il faut préciser que malgré le caractère obligatoire, ces simulations de crise ne

sont en aucun cas utilisés pour calculer l'exigence en fonds propres. Cependant, le but de ces simulations est de compléter le dispositif de mesure des risques.

Par ailleurs, en réponse à l'évolution remarquable et au développement notable des activités bancaires, les autorités réglementaires ont décidé d'imposer un certain nombre de normes, en matière de contrôle des risques bancaires et en particulier du risque de marché. Ces règlements sont inspirés des recommandations du Comité de Bâle, groupe de travail institué en 1975 par les gouverneurs des banques centrales, qui tente d'établir des normes minimales de contrôle bancaire, communes à tous les pays.

Les réglementations prudentielles imposant aux banques la détention de capitaux propres minimaux reposent sur l'idée, que la banque est en situation de faillite si elle n'est définitivement plus en mesure de faire face à ses engagements, notamment parce qu'elle préfère exercer sa clause de responsabilité limitée, les actionnaires refusent de lui apporter les fonds nécessaires à la poursuite de ces activités. Les actionnaires estiment alors que la totalité des encaissements que les actifs sont susceptibles de produire ne permet pas de couvrir la totalité des décaissements associés aux passifs.

Il faut bien signaler que les normes de contrôle ne visent pas, en fin de compte, à faire disparaître les risques, mais à les mesurer afin de les maîtriser et à couvrir, ceux jugés les plus importants. En pratique, ces normes s'expriment essentiellement comme une exigence pour la banque d'un montant minimal de fonds propres réglementaires, destinés à la couverture des risques de marché. Quant à l'évaluation, le comité de Bâle, dans le cadre de la réglementation prudentielle, laisse la liberté aux institutions de choisir leur modèle de gestion interne. Il serait donc intéressant de voir comment modéliser et prévoir le risque par la meilleure méthode. En revanche, les institutions doivent respecter un certain nombre de règles de gestion, tel que le calcul quotidien de la perte potentielle maximale $P(t)$ appelée aussi Value at Risk (VaR), et ce pour une période de détention de dix jours (l'horizon T) et pour un niveau de confiance de 99%. Dans ce sens, nombreuses méthodes ont été développées et consacrées au calcul et à l'estimation de la VaR.

Parmi les approches les plus utilisées, nous pouvons mentionner la VaR historique qui consiste à obtenir sur la base d'un historique des variations des facteurs de risque, une loi de probabilité relative aux variations des valeurs du portefeuille. De cette loi, on peut extraire un quantile qui permet d'identifier la VaR pour un seuil de confiance donné. La VaR Monte Carlo, quant à elle, est

basée sur la variation de la VaR historique dans le cas où, seule la matrice de variance-covariance est disponible. Ainsi, cette méthode consiste à générer par une simulation de Monte Carlo, des scénarios compatibles avec cette matrice. La VaR paramétrique (ou analytique) est une troisième méthode qui consiste à représenter le gain d'un portefeuille comme combinaison linéaire des facteurs gaussiens.

Vu qu'elle se place explicitement ou implicitement dans le cadre de l'approximation par la loi normale, la VaR apparaît inadaptée à la prise en compte des événements extrêmes, et qui sont justement ceux qui font peser les risques les plus grands dans le monde des affaires et des banques. En plus, dans un contexte de gestion des risques, la loi gaussienne peut conduire à une sous-évaluation importante. D'ailleurs, depuis l'étude de Mandelbrot (1963) sur les prix du coton, la présence des "queues épaisses" dans les lois des séries financières est un fait empirique très connu.

C'est dans ce cadre, que la théorie des valeurs extrêmes développée depuis 1928 par Fisher et Tippett, a récemment trouvé sa place dans le domaine de la finance et des banques afin d'étudier les fluctuations des cours extrêmes. Ainsi, divers travaux ont été menés, notamment ceux de Müller et al. (1998, [101]), et Danielsson et De Vries (1997, [29]).

4 Application of POT approach to CAC40 return index

4.1 Introduction

Le rendement représente la variable la plus communément étudiée en finance. Cependant, dans la quasi totalité des cas traités, les rendements sont supposés normaux et ils sont traités en logarithme pour éliminer les problèmes de non stationnarité, souvent présents dans les données financières.

Même si l'hypothèse forte, de normalité ou lognormalité des rendements, continue à être présente dans la plupart des travaux actuels en finance, Mandelbrot a montré depuis 1963 que les lois de certains actifs sur des périodes bien définies, ne s'identifient pas à des lois gaussiennes. En outre, les lois correspondantes ont un comportement asymptotique plus lent que celui de la loi normale. A l'évidence, même la volatilité affectant les marchés financiers, varie au cours du temps d'une façon asymétrique, où en général, les baisses et les hausses

sont fréquemment suivies des variations de même sens (Engle (1982, [46]), Bollerslev et al. , (1992, [10])). Ainsi, il est tout à fait légitime de se demander comment prendre en compte les queues des lois plus épaisses que celles d'une gaussienne et dont les variations, sont fondamentalement liées à l'existence des événements rares ou extrêmes.

Dans le même ordre d'idées, nous étudions dans cette partie, les modèles des valeurs extrêmes qui permettent de mieux caractériser les rendements de CAC40. Cet indice présente une grande importance dans le monde de la finance, et particulièrement dans la bourse de Paris.

Les modèles des valeurs extrêmes qui vont être appliqués sont basés sur l'approche POT. La difficulté de cette approche, comme il a été bien signalé pour les données hydrologiques, réside dans l'identification d'un seuil optimal. Ainsi, nous présentons en premier point, le traitement de la série CAC40 pour adapter les données aux conditions d'application de l'approche POT. En deuxième point, nous comparons les performances de certaines approches de détection du seuil optimal, notamment l'approche ATSM, MTM et MSE.

4.2 Données

Le CAC40 est né officiellement le 15 Juin 1988, juste après le crash de 1987. C'est un indice qui désigne une référence du marché boursier français, vu qu'il représente une mesure pondérée de la capitalisation des 40 plus importantes valeurs, parmi les 100 valeurs les plus élevées dans la Bourse de Paris (Euronext Paris) et par conséquent, il permet de donner une idée sur la tendance globale du marché.

L'histoire du CAC40 permet de constater qu'il a connu plusieurs changements depuis sa création. Nous pouvons évoquer à titre d'exemple que l'indice a atteint les 6944.77; son point culminant, le 4 Septembre 2000 pour rechuter suite au crash boursier de 2008 et perdre plus de 43.5 % de sa valeur.

L'analyse des 5222 valeurs mensuelles du CAC40 de la période allant du 3 Mars 1990 au 20 Décembre 2010, permet de mettre en exergue la forte volatilité de l'indice.

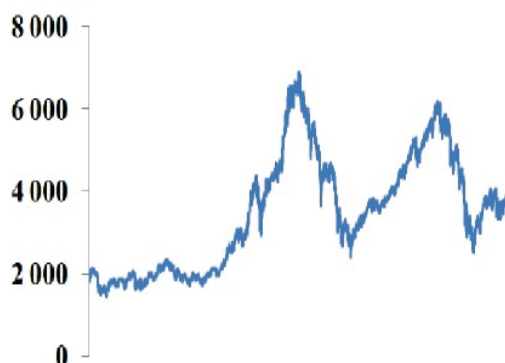


FIGURE 6.1 – Evolution du CAC40 durant la période allant du 3 Mars 1990 au 20 Décembre 2010

En effet, cette période a été spécialement caractérisée par un certain nombre d'événements qui ont marqué l'histoire du CAC40. Nous pouvons citer la crise énergétique d'août 1990, la crise russe en septembre 1998, la bulle d'internet en fin 1999/2000, les attentats du 11 septembre 2001 et le début de la guerre d'Irak en mars 2003.

Sur le plan pratique, la forte volatilité du CAC40 expose les investisseurs à plusieurs risques. En plus, en ce qui concerne l'aspect de la modélisation, elle met en doute l'hypothèse de la stationnarité et de la normalité. Ces faits ensemble, justifient bien l'intérêt porté à la modélisation des lois de probabilité relative au CAC40. Nous suggérons une approche de modélisation basée sur POT. Pour ce faire, nous devons vérifier d'abord les hypothèses d'application, notamment celle qui assume que les observations traitées sont i.i.d. Vu que la loi est inconnue, nous proposons d'utiliser les tests non paramétriques d'indépendance et d'homogénéité.

L'application du test de Mann-Whitney montre que les données d'origine (les cours d'indice CAC40) ne vérifient pas la condition d'indépendance et d'homogénéité. Afin de répondre aux exigences théoriques, nous considérons l'inverse du rendement du CAC40, qui vérifie bien, selon les mêmes critères, la condition i.i.d.

Par ailleurs, le calcul des coefficients d'asymétrie et du kurtosis montrent que la loi des données transformées ; l'inverse du rendement du CAC40, est étalée vers la droite. En effet, le coefficient d'asymétrie est égal à 24.42 et le coefficient du kurtosis est supérieur à 3. Ces constatations confirment que la loi des données transformées est positivement asymétrique. Nous pouvons conjecturer dès lors, que la loi des excès correspondant à l'inverse du rendement du CAC40, peut être bien ajustée à une loi de Fréchet.

4.3 Méthodologies

Le choix du seuil constitue une étape cruciale dans l'approche POT. Comme il a été signalé auparavant, plusieurs approches ont été développées pour identifier le "bon" seuil. Nous proposons dans cette partie la comparaison des approches SEM et MTM précédemment présentées, avec une troisième méthode appelée Automated Threshold Selection Method (ATSM) et développée par Thompson et al. (2009, [143]). Le fondement théorique d'ATSM est bien détaillé dans le même article de Thompson et al. Nous nous contentons dans ce qui suit de mettre en exergue l'algorithme récapitulant les étapes d'application de l'approche ATSM :

Etape 1 : Identifier un intervalle de seuils candidats : $u_1 < u_2 < \dots < u_n$.

Pour minimiser le temps de calcul, les valeurs médiane et le 0.98 quantile pourront être affectées à u_1 et u_n respectivement. A signaler que le seuil u_1 doit être suffisamment grand pour assurer une estimation fiable.

Une fois l'intervalle des seuils candidats est identifié, calculer pour $j = 1, \dots, n$ les estimateurs du paramètre d'échelle $\hat{\sigma}_{u_j}$ et de forme $\hat{\xi}_{u_j}$, obtenus à partir des dépassements au delà du seuil u_j .

Etape 2 Si u est le seuil optimal, alors selon Thompson et al. (2009, [143]), pour tout $\tau_{(u_j)} = \hat{\sigma}_{u_j} - \hat{\xi}_{u_j} u_j$, u_ν et $u_{\nu-1}$ avec $u_\nu \geq u_{\nu-1}$, la différence $\tau_{(u_\nu)} - \tau_{(u_{\nu-1})}$ suit asymptotiquement une loi normale de moyenne 0.

Considérer alors $j = 2$ et tester l'hypothèse $H_0 : (\tau_{(u_\nu)} - \tau_{(u_{\nu-1})})_{j \leq \nu \leq n}$ est distribuée selon une loi normale de moyenne 0". Si H_0 est acceptée, alors $u_{j-1=1}$ est le seuil optimal. Sinon, considérer $j = 3$ au lieu de $j = 2$ et retester H_0 . Répéter cette procédure avec les seuils candidats de l'intervalle considéré au départ, jusqu'à l'obtention du seuil optimal ou $j = n$.

Selon Thompson et al. (2009, [143]), l'algorithme mentionné ci-dessus peut ne pas converger. Cependant, ce cas se produit très rarement.

4.4 Résultats

Nous utilisons MRL plot pour identifier l'intervalle des seuils candidats, que nous devons utiliser pour initialiser les étapes d'application des méthodes MTM, SEM et ATSM. Ce graphique est approximativement linéaire sur l'intervalle des seuils $[0; 3000]$. Pour les seuils supérieurs à 3000, le graphique montre beaucoup d'instabilité. Cela est dû à la petite taille de l'échantillon considéré (de 5222 observations initiales, seulement 64 observations qui subsistent

4. APPLICATION OF POT APPROACH TO CAC40 RETURN INDEX 95

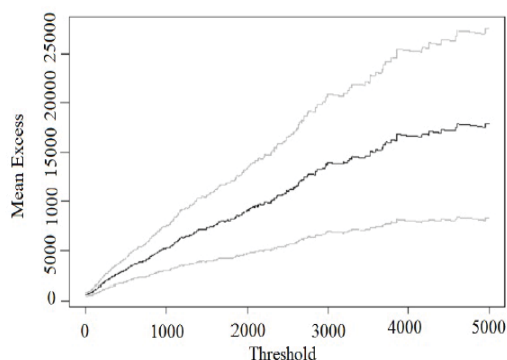


FIGURE 6.2 – MRL Plot de l'inverse du rendement de CAC40

TABLE 6.1 – Lois ajustées selon les méthodes retenues et leurs p-values d'AD2up

Approche	Méthode d'estimation	Seuil optimal	p-value
MTM	Moment	1000	0.14
MTM	Maximum de vraisemblance	1156	0.06
SEM	Moment	713	$< 2.210^{-6}$

au delà du seuil 3000). Nous discrétisons l'intervalle $[0; 3000]$ en considérant $\{u_0 = 0, u_1 = \Delta, u_2 = 2\Delta, u_3 = 3\Delta, \dots, u_n = n\Delta = 3000\}$ $\Delta = 0.1, 0.01, 0.001$

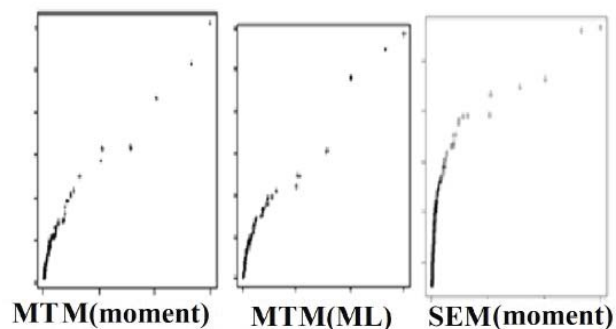


FIGURE 6.3 – QQ plot des différentes lois ajustées

que les excès de l'inverse du rendement du CAC40 s'ajuste à une loi à queue lourde.

Ainsi, la loi la plus appropriée aux excès de l'inverse du rendement du CAC40 est donnée par :

$$F_{1000}(x) = F_{1000}(x, 2443.19, 0.45) = 1 - \left(1 + 0.45 \frac{x - 1000}{2443.19}\right)^{-\frac{1}{0.45}} \quad (4.1)$$

Il s'agit d'une loi de Fréchet, permettant de caractériser la forte volatilité de l'indice considéré et le risque des pertes, et par conséquent, d'éclairer les décisions qui portent sur des investissements liés à l'indice CAC40.

Conclusion Générale

La complexité des marchés boursiers et les caractéristiques stochastiques des données hydrologiques, ont conduit les chercheurs à s'intéresser à l'étude des extrêmes qui relèvent du domaine de la bourse ou de l'hydrologie. L'un des aspects importants lors de l'étude de ce type de données, réside dans la recherche du rendement maximal d'une action ou le niveau maximal de la crue. En effet, leur estimation est très importante pour prévoir les niveaux et les périodes de retour, et par conséquent pour prendre les bonnes décisions de gestion et d'investissement, ou d'orienter des plans d'actions et d'aménagement.

Dans ce travail de recherche, nous nous sommes intéressés aux méthodes statistiques utilisées pour la réalisation de cet objectif. La théorie des valeurs extrêmes est manifestement l'approche la plus utilisée. Cependant, les résultats théoriques de cette approche sont souvent confrontés dans la pratique à l'invalidité des conditions d'application. Ce travail avait donc pour objectif de mettre en application la théorie des valeurs extrêmes, pour le traitement des données qui relèvent des domaines de la bourse et de l'hydrologie, en utilisant l'approche POT qui constitue une alternative à l'approche GEV, car en terme de biais et de variance, elle conduit à des meilleurs résultats. Cependant, la difficulté de cette méthode réside dans le choix du seuil. De ce fait, nous nous sommes focalisés beaucoup plus dans ce travail, sur la comparaison des méthodes de détection du seuil optimal; étape cruciale pour un bon ajustement de la loi GPD, ainsi que sur l'adaptation de la théorie aux données qui présentent généralement des caractères de dépendance, de non stationnarité ou d'hétérogénéité. Selon nos applications mises en œuvre dans cette thèse, nous pouvons privilégier le choix de la méthode LRT et MTM par rapport aux autres méthodes.

A signaler que même si les simulations permettent de répondre à un certain nombre de questions qui portent sur le choix des tailles des échantillons, des lois empiriques et des méthodes d'estimation, nous avons préféré lors du présent travail de recherche, de traiter des données réelles au lieu des données simulées, qui ne reflètent jamais et d'une façon intime la complexité des observations concrètes. Pour ce faire, un recueil bibliographique était indispensable avant d'entamer le développement théorique nécessaire à la mise en œuvre des

méthodes choisies. Nous avons évoqué donc les intérêts des décideurs et des investisseurs quant à la gestion des risques hydrologiques et boursiers et nous avons enchaîné ensuite, par la présentation des effets et des impacts des phénomènes extrêmes illustrés par des crises, des crashes ou des inondations. En outre, nous avons fait référence à un recueil de pratiques généralement utilisées pour la modélisation des risques et nous avons montré, la suprématie des approches qui relèvent de la théorie des valeurs extrêmes.

Pour les données hydrologiques ou les données boursières, les résultats de notre thèse confirment que la meilleure loi qui peut ajuster les excès utilisés dans cette thèse, est celle de Fréchet. Cette loi à queue lourde permet de caractériser les données par une forte chance d'atteindre des niveaux élevés. Ce fait permet de mettre l'accent sur la nécessité d'une prise d'attention du côté des décideurs ou d'investisseurs, quant à la fréquence des événements extrêmes qui peuvent causer des dommages et des dégâts matériels et humains.

De part l'exploration des méthodes subjectives destinées à la détection du seuil optimal, la comparaison de leurs résultats disséminés à l'occasion du Second Workshop on the Application of Stochastic Models in Finance and in Risks Management (Rabat, 2012) et 5th International Perspective on Water Resources and the Environment Conference (Marrakech, 2012) ainsi que la publication des deux articles dans *Journal of Hydrologic Engineering*, 2014 (Managing hydrological risks with extreme modeling : application of peaks over threshold model to the Loukkos watershed) et *International Journal of Statistics and Economics*, 2012 (Statistical methods to expect extreme values : Application of POT, approach to CAC40 return index), la présente thèse a permis en plus, d'explorer des pistes de recherches, dont une part de résultats est inachevée et par conséquent, elle ne figure pas dans le présent manuscrit, et dont l'autre part est encore à tester et à valider.

Pour conclure, nous signalons que notre thèse n'est qu'un point de départ pour mieux connaître ce monde immense des valeurs extrêmes. Nous projetant dans l'avenir d'explorer d'autres pistes de recherche. Sur le plan théorique, nous envisageons d'une part, de comparer les résultats des modèles GEV non stationnaires avec la technique des copules conditionnelles et la régression de quantiles, et de combiner d'autre part, les copules avec l'approche POT.

Sur le plan purement expérimental, nous souhaitons confronter nos méthodes avec plusieurs approches spécifiques telles que les méthodes intégrées dans les logiciels spécialisés, et connus dans le domaine de la bourse ou de l'hydrologie. Nous souhaitons élargir nos applications et s'intéresser aux risques qui

touchent d'autres domaines, tels que les risques relevant de l'actuariat, de la démographie ou de l'économie. Nous sommes actuellement en train d'explorer l'efficacité de la TVE dans le domaine bancaire, notamment en ce qui concerne le calcul de la VaR et la détermination des fonds propres exigés pour se prémunir contre les crises. Parallèlement à ce travail, une autre étude est encore en cours d'achèvement. Elle porte sur la contribution des copules pour la gestion des risques en exploitant la relation entre inflation et taux de change.

Bibliographie

- [1] Atyeo J. and Walshaw D., *A hierarchical model for extreme rainfall incorporating regional variation, temporal trend and local dependence. Applied Statistics*, 2011.
- [2] Beirlant J., Goegebeur Y., Teugels J. and Segers J., *Statistics of extremes. Wiley Series in Probability and Statistics. John Wiley and Sons Ltd*, Chichester, 2004.
- [3] Beirlant J., Dierckx G., Goegebeur Y. and Matthys G., *Tail index estimation and an exponential regression model. Extremes*, 2, 177-200.
- [4] Beirlant J., Teugels J. and Vynckier P., *Practical Analysis of Extreme Values, Leuven University Press*, 1996.
- [5] Benkhaled A., *Statistical distributions of annual maximum rainfalls depths in the area of cheliff. comparison of techniques and results, Courrier du Savoir*, 08 , pp.83-91, 2007.
- [6] Berman S.M., *Asymptotic Independence of the Numbers of High and Low Level Crossings of Stationary Gaussian Processes. Ann. Math. Statis.*, 42, 927-945, 1971.
- [7] Berman S.M., *Limit Theorems for the Maximum Term in Stationary Sequences. Ann. Math. Statis.*, 35, 502-516, 1964.
- [8] Bernard-Michel C., Gardes L., Girard S. and Molinié G., *Première analyse des pluies extrêmes dans la région Cévennes-Vivarais, INRIA*, October, 2008.
- [9] Blanchet J. and Lehning M., *Mapping snow depth return levels : Smooth spatial modeling versus station interpolation. Hydrology and Earth System Sciences*, 14 :2527-2544, 2010.
- [10] Bollerslev T., Chou R.Y. and Kroner K.F., *ARCH modeling in finance : a selective review of the theory and empirical evidence. Journal of Econometrics* 52, p.5-59, 1992.

- [11] Borchani A., *Statistiques des valeurs extrêmes dans le cas de lois discrètes*, ESSEC Working Paper 10009, 2010.
- [12] Bortot P., Coles S.G. and Tawn J.A. *The multivariate gaussian tail model : an application to oceanographic data. Applied Statistics*, 49 :31-49, 2000.
- [13] Bouyé E., Nikeghbali A., Riboulet G. and Roncalli T. *Copulas for finance- A reading guide and some applications*, Groupe de Recherche Opérationnelle, Crédit Lyonnais, Working Paper, 2000.
- [14] Breiman L., Stone C.J. and Gins J.D. *New methods for estimating tail probabilities and extreme value distribution.*, Tech Report TSC-PD-A226-1, Technology Service Corporation, Santa Monica.
- [15] Brunet-Moret Y., *Homogénéisation des précipitations In : Cah» Orstom, sér. Hydrol.*, vol XVI, 3 et 4, p. 147-170, 1979.
- [16] Bruun J.T and Tawn J.A, *Comparison of approaches for estimating the probability of coastal flooding. Applied Statistics*, 47 :405-423, 1998.
- [17] Capéraà P., Fougères A.L. and Genest C., *CEstimation of a bivariate extreme value distribution, Extremes*, 3, 311-329, 2000.
- [18] Charlot F. and Merad D., *Sur les méthodes de processus ponctuels et de renouvellement dans les systèmes de files d'attente. III- Processus ponctuels périodiques et files d'attente périodiques. Annales de la Faculté des sciences de l'Université de Clermont, Série Probabilités et applications*, no. 8, p. 63-77, 1989.
- [19] Cherubini U. and Luciano E., *Multivariate option pricing with copulas*, International Centre for Economic Research in its series ICER Working Papers, Applied Mathematics Series with number 05, 2000.
- [20] Chernoff H., *On the Distribution of the Likelihood Ratio*, Ann. Math. Statist. Volume 25, Number 3, p. 573-578, 1954.
- [21] Christidis N., Stott P.A. and Brown S.J., *The role of human activity in the recent warming of extremely warm daytime temperatures, Journal of Climate, Vol 24, p. 1922-1930*, 2011.
- [22] Christopeit N., *Estimating parameters of an extreme value loi by the method of moments, Journal of Statistical Planning and Inference 41, p. 173-186*, 1994.
- [23] Coles S.G., *An Introduction to Statistical Modeling of Extreme Values*, Springer, 2001.

- [24] Coles S.G. and Tawn J.A. *Modelling extreme multivariate events*, *Journal of the Royal Statistical Society*, B 53 :377-392, 1991.
- [25] Coles S.G. and Tawn J.A. *Statistical methods for multivariate extremes : an application to structural design (with discussion)*. *Applied Statistics*, 43 :1-48, 1994.
- [26] Coles S.G. and Tawn J.A. *A bayesian analysis of extreme rainfall data*. *Applied Statistics*, 45 :463-478, 1996.
- [27] Dahan E. and Mendelson H., *An Extreme-Value Model of Concept Testing*, *Management Science*, 2001.
- [28] Danielsson J., De Haan L. and De Vries C. G., *Using a bootstrap method to choose the sample fraction in tail index estimation*, *Journal of Multivariate Analysis*,76, 2001, p. 226-248.
- [29] Danielsson J. and De Vries C. G., *Tail index and quantile estimation with very high frequency data*. *Journal of Empirical Finance*,4, 1997, p. 241-257.
- [30] Deidda R., *A multiple threshold method for fitting the generalized Pareto distribution to rainfall time series*, *Hydrology and Earth System Sciences*,14, 2010, p. 2559-2575.
- [31] Delmas J.F and Jourdain B., *Modèles aléatoires, Applications aux sciences de l'ingénieur et du vivant*. Springer,2006.
- [32] Draisma, G., de Haan, L., Peng,L. and Pereira, T.T, *bootstrap based method to achieve optimality in estimating the extreme value index*. *Extremes*, 2, 1999, p.367-404.
- [33] Davison A.C. and Smith R.L., *Models for exceedances over high thresholds* *J.R. Statist. Soc.*, B 52, 3, pp.393-442, 2001.
- [34] de Haan L. and Mendelson H., *A spectral representation for max-stable processes*. *Annals of Probability*, 12 :1194-1204, 1984.
- [35] de Haan L. and Rootzen H., *On the estimation of high quantiles* *J. of Statistical Planning and Inference*,35, n.1, 1-13, 1993.
- [36] Deheuvels P., *Strong Approximation in Extreme Value Theory and Applications*, *Colloquium on Limit Theorems in Probability and Statistics*, North-Holland,1, 1984.
- [37] Deheuvels P., *La fonction de dépendance empirique et ses propriétés - Un test non paramétrique d'indépendance*, *Académie Royale de Belgique - Bulletin de la Classe des Sciences - 5e Série*, 65, 274-292, 1979.

- [38] Dekkers A., Einmahl J. and De Haan L., *A moment estimator for the index of an extreme-value loi*, *Ann. Statist.* 17, p. 1833-1855, 1989.
- [39] Drees H., *Refined Pickands Estimators of Extreme Value Index*, *Ann. Statist.*, 23, p. 2059-2080, 1995.
- [40] El Adlouni S. and Ouarda B.M.J., *Frequency Analysis of Extreme Rainfall Events, Chapitre du livre (Rainfall : State of the Science), American Geophysical Union (AGU). Geophysical Monograph Series., VOL. 191, 287 PP, 2010.*
- [41] El Adlouni S., Bobée B., Ouarda B.M.J., *Caractérisation des distributions à queue lourde. Rapport de Recherche N° R-929, Hydro-Québec/CRSNG, 2007.*
- [42] El Adlouni S. and Ouarda B.M.J., *Comparaison des méthodes d'estimation des paramètres du modèle GEV non stationnaire*, *Journal of Water Science.*, 21, 1, p. 35-50, 2008.
- [43] Einmahl J.H.J. and Smeets S.G.W.R., *Ultimate 100m world records through extreme-value theory. Statistica Neerlandica, 65(1), 32-42, 2011.*
- [44] Embrechts P., Kluppelberg C. and Mikosch T., *Modelling extremal events, Springer Verlag, Berlin, 1997.*
- [45] Embrechts P., McNeil A.J. and Straumann D., *Correlation : pitfalls and alternatives. RISK, 69-71, 1999.*
- [46] Engle R.F., *Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, p.987-1007, 1982.*
- [47] Fama E., *The behavior of stock prices. Journal of Business, 38, p. 34-105, 1965.*
- [48] Fougères A.L., *Multivariate extremes, In Extreme Values in Finance, Telecommunications, and the Environment. B. Finkenstädt and H. Rootzén (eds), Boca Raton, FL : Chapman & Hall/CRC., 2004.*
- [49] Galambos J., *The asymptotic theory of Extreme Order Statistics, Wiley, New-York, 1985.*
- [50] Genest C. and MacKay R.J. *Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. La rev. canad. statist., 14, 145-159, 1986.*
- [51] Greenwood J., Landweher J., Matalas N. and Wallis J., *Probability weighted moments : Definition and relation to parameters of several lois expressible in inverse form, Water Resources Research 15, p. 1049-1054, 1979.*

- [52] Garavaglia F., *Méthode SCHADEx de prédétermination des crues extrêmes*, Université Joseph Fourier de Grenoble, Grenoble, 389 p, 2011.
- [53] Garrido-Charras M., *prédiction des événements rares et estimation des quantiles extrêmes, méthodes de sélection de modèles pour les queues de distribution*, Université J. Fourier, 2002.
- [54] Gnedenko B., *Sur la distribution limite du terme maximum d'une série aléatoire*, Ann. Math., 44(3), 423-453, 1943.
- [55] Gomes M.I. et Oliveira O. , *The bootstrap methodology in statistics of extremes : theory and applications - choice of the optimal sample fraction. Extremes*, 4, p.331-358.
- [56] Guillou A. and Willems P., *Application de la théorie des valeurs extrêmes en hydrologie*, Rev. Statistique Appliquée, 2006, LIV (2), 5-31, 2006.
- [57] Guillou A., Naveau P., Diebolt J. and Ribereau P. *Return level bounds for small and moderate sample sizes for discret and continuous random variables*, 2007.
- [58] Gumbel E.J., *Statistics of Extremes*, Columbia University Press, New-York, 1958.
- [59] Gumbel E.J. and Mustafi C.K., *Some analytical properties of bivariate extremal distributions*, Am. Stat. Assoc. J., 62, p.569-589, 1967.
- [60] Hill B. M., *A simple general approach to inference about the tail of a Loi*, Annals of Statistics 5, p. 1163-1174, 1975.
- [61] Hols M.C.A.B. and de Vries C.G., *The limiting distribution of extremal exchange rate returns*, Journal of applied econometrics, p. 287-302, 1991.
- [62] Horiuchi S. and Wilmoth J.R., *Deceleration in the age pattern of mortality at older ages*, Demography, 35(4) : 391-412, 1998.
- [63] Hosking J.R.M., Wallis J.R. and Wood E.F., *Estimation of the Generalized Extreme Value distribution by the method of probability weighted moments*, Technometrics. 27,p. 251-261, 1985.
- [64] Hougaard P., *A class of multivariate failure time distributions*, Biometrika. 73,p. 671-678, 1986.
- [65] Hsing T., *Point Process Associated with Extreme Value Theory*, Technical Report, University of North Carolina, Chapel Hill, 1984.
- [66] Hubert P., Bendjoudi H., *Introduction à l'étude des longues séries pluviométriques. In : Journées hydrologiques de l'ORSTOM*, , Montpellier, pp. 20, 1996.

- [67] Hüsler J., *Maxima of Bivariate Random Vectors : Between Independence and Complete Dependence Statistics and Probability Letters*, 21, pp. 385-394,, 1994.
- [68] Hüsler J., *Extreme Values and Rare Events of Non-Stationary Random Sequences Dependence in Probability and Statistics : A Survey of Some Recent Results E.Eberlein, M.S. Taquq ed*, 1985.
- [69] Ibrajimov I.A., Linnik Yu.K., *Independant and Stationary Sequences of Random Variables, Walters-Noordho Publi., Groningen.*, 1971.
- [70] Jansen D.W. and de Vries C.G. *On the frequency of large stock returns : putting booms and busts into perspective, review of Economics and Statistics*, 73, p.18-24 1991.
- [71] Jenkinson A.F., *The frequency distribution of the annual maximum (or minimum) values of meteorological events, Journal of the Royal Meteorological Society* , 81, p. 158-272, 1955.
- [72] Joe H. and Xu J.J., *The estimation method of inference functions for margins for multivariate models." Technical Report no. 166, Department of Statistics, University of British Columbia*, 1996.
- [73] Joe H., *Multivariate models and dependence concepts, Chapman et Hall*, 1997.
- [74] Jondeau E. and Rockinger M. *Extreme value on stocks markets : the tails of densities on G5 and emerging financial markets, Journée de la Fondation Banque de France* , Octobre, 1999.
- [75] Kawas M.L. and Moreira R.G., *Characterization of prosect quality attributes of tortilla chips during the frying process. Journal of Food Engineering*, 47, 97-107, 2001.
- [76] Katz R.W, *Techniques for estimating Uncertainty in climate change Scenarios and impact studies Climate research*, 20, 167-185, 2002.
- [77] Khaliq M.N., Ouarda T.B.M.J., Ondo J.,-C, Gachon P. and Bobée B., *Frequency analysis of a sequence of dependent and/or non-stationary hydrometeorological observations : a review. Journal of hydrology*, 329 :534-552, 2006.
- [78] Khoudraji A., *Contribution à l'étude des copules et à la modélisation de valeurs extrêmes multivariées*, Thèse de doctorat, Université Laval, Québec, 1998.
- [79] Klajnmic H., *Niveaux de retour pour les vitesses extrêmes des vents, XXXV èmes Journées de Statistique, Lyon*, 2003.

- [80] Kratz M. and Resnick S. *The qq-estimator and heavy tails*, *Commun. Statist. Stochastic Models*, 12, 699-724, 1996.
- [81] Lai T.L. and Robbins H., *Maximally Dependent Random Variables*, *Proc. Nat. Acad. Sci. U.S.A.*, 73, 286-288, 1976.
- [82] Lang M., Renard B., Kochanek K., Arnaud P., Aubert Y. *Projet Extra-Flo : comparaison de différents cadres d'analyse (locale, régionale, locale-régionale) pour l'estimation de quantiles de crue*, 2012.
- [83] Leadbetter R., Lindgren G. and Rootzen H., *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag. N.Y. p. 29-63, 1983.
- [84] Lekina A., *Estimation non paramétrique des quantiles extrêmes conditionnels*, Thèse de doctorat sous la direction de Stéphane Girard & Laurent Gardes, INRS, Québec, Canada., 2010.
- [85] Lévy P., *Théorie des erreurs. La loi de Gauss et les lois exceptionnelles*, *Bulletin de la Société Mathématique de France*, 52, p. 49-85, 1924.
- [86] Li D.X., *On Default Correlation : A Copula Function Approach*, RISK-METRICS, Working Paper Number 99-07, 2000.
- [87] Longin F., *The choice of the distribution of asset returns : How extreme value theory can help ?*, *Journal of Banking and Finance*, 29, 2005.
- [88] Longin F., *From value at risk to stress testing : The extreme value approach*, *Journal of Banking and Finance*, 24, p.1097-1130, 2000.
- [89] Longin F., *The asymptotic distribution of extreme stock market returns*, *Journal of Business*, 69, p.383-408, 1996.
- [90] Longin F., *Value at risk : Une nouvelle approche fondée sur les valeurs extrêmes*, *Annales d'économie et de statistique*, 52, p. 23-51, 1998.
- [91] Longin F., *Booms and Crashes. Application of Extreme Value Theory to the U.S. Stock Market*, Institute of Finance and Accounting, London Business School. IFA Working Paper p.179-93, 1993.
- [92] Loretan M. and Philipps P.C.B. *Testing the covariance stationnarity of heavy-tailed time series*, *Journal of Empirical Finance*, 1, p. 211-248, 1994.
- [93] Loynes R.M., *Extreme Values in Uniformly Mixing Stationary Stochastic Processes*, *Ann. Math. Statist.*, 36, p. 993-999, 1965.
- [94] Lubes H. and Masson J., *Méthode des moments de probabilité pondérés. Application à la loi de Jenkinson*, *Hydrol. continent.*, vol. 6, n 1, p. 67-84, 1991.

- [95] Lye L., Hapuarachchi K. et Ryan S., *Bayes estimation of the extreme-value reliability function*, *IEEE Transactions on Reliability* 42, p. 641-644, 1993.
- [96] Marshall A.W and Olkin I., *Families of multivariate distributions*, *JASA*, Vol 83, n 403, p. 834-841, 1988.
- [97] Mandelbrot B.B., *The Variation of Certain Speculative Prices*, *Journal of Business*, 26, p. 394-419, 1963.
- [98] Mc Neil A.J. and Frey R., *Estimation of tail-related risk measures for heteroscedastic financial time series : An extreme value approach* *Journal of Empirical Finance*, 7, p. 271-300, 2002.
- [99] Mélou M., *VaR : quelle mise sur la table du casino ?*, *Les Echos*, 2008.
- [100] Morton I.D., Bowers J. and Mould G. *Estimating return period wave heights and wind speeds using a seasonal point process model*. *Coastal Engineering*, 31 :305-326, 1997.
- [101] Müller U.A., Dacorogna M.M. and Pictet O.V. *Heavy tails in high-frequency financial data*, *Birkhäuser*, Boston, p.55-77, 1998.
- [102] Nikias C.L. and Shao M. *Signal Processing with Alpha-Stable Distributions and Applications*, New-York, Wiley-Interscience, 1995.
- [103] Nolan J., *Numerical approximation of stable densities and distribution functions*, Preprint American University Washington, 1996.
- [104] Parisi F. and Lund R., *Seasonality and return periods of landfalling atlantic basin hurricanes*. *Australian and New Zealand Journal of Statistics*, 42(3) :271-282 , 2000.
- [105] Pereira V., Giremus A. and Grivel E., *Modeling of Multipath Environment Using Copulas for Particle Filtering Based GPS Navigation*. *IEEE Signal Process. Lett.*, 19(6) : 360-363, 2012.
- [106] Philipp M., *Mixing Sequences of Random Variables and Probabilistic Number Theory*, *Memoirs. Amer. Math. Soc.*, 114 , 1971.
- [107] Pickands J., *Statistical inference using extreme order statistics*, *Annals of Statistics* 3, p. 119-131, 1975.
- [108] Pickands J., *Multivariate extreme value distributions*. *Bull. int. Statistics.Ins*, 1981.
- [109] Pratt J.W and Zeckhauser R.J , *Proper Risk Aversion,* " *Econometrica*, *Econometric Society*, vol. 55(1), p.143-54, 1987.

- [110] Prescott P. and Walden A.T., *Maximum likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution for censored samples*, *Journal of Statistical Computation and Simulation*, 16(3-4) :241-250, 1983.
- [111] Prescott P. and Walden A.T., *Maximum likelihood estimation of the parameters of the generalized extreme-value distribution*, *Biometrika*, 67, 723-724, 1980.
- [112] Prevosto M., *Comparaison de méthodes de simulation non linéaire de la surface libre en vagues irrégulières pour la prédiction de hauteurs extrêmes*, *Ifremer, TMSI/IDM/COM 03-0H1*, 2003.
- [113] Raillard N., *Comportements extrêmes de processus spatio temporels. Application en océanographie et météorologie*, *Thèse de doctorat réalisée sous la direction de MM. Jian feng YAO et Bertrand CHAPRON*, Université de Rennes 1, 2011.
- [114] Raggad B., *Fondements de la théorie des valeurs extrêmes, ses principales applications et son apport à la gestion des risques du marché pétrolier*, *Mathematics and Social Sciences*. p. 29-63, 2009.
- [115] Reiss R.D. and Thomas M., *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, *Birkhauser Basel*. 2nd edition, 2001.
- [116] Reiss R.D. and Thomas M., *Statistical Analysis of Extreme Values*, *Birkhauser Verlag, Boston, MA*, 1997.
- [117] Resnick S.I., *Extreme Values, Regular variation and Point Processes*, *Springer-Verlag, New-York*, 1987.
- [118] Resnick S.I., *C. Starica Consistency of Hill's Estimator for Dependent Data*, *J. Appl. Probab.*, 32, p. 139-167, 1987.
- [119] Ribereau P., Toulemonde G. and Gannoun A., *Théorie des Valeurs Extrêmes en Sciences du Climat*, Université Montpellier 2, 2010.
- [120] Rivoirard J., *De l'analyse structurale à la modélisation des ressources naturelles : contributions géostatistiques, thèse, Université Pierre et Marie Curie-Paris 6*, 2012.
- [121] Roberts S.J., *Extreme value statistics for novelty detection in biomedical data processing. IEEE proceedings, Science Measurement and Technology*, 147, 2000.
- [122] Roncalli T., *Modélisation des Evénements Rares pour la Gestion des Risques*, DESS 203 de l'Université Paris IX Dauphine, 2002.

- [123] Rosbjerg D., *On the annual maximum distribution in dependent partial duration séries. Stochastic Hydrol. Hydraul.*, 1, 3-16, 1987.
- [124] Rosenblatt M., *Remarks on some non parametric estimates of a density function. Ann. Math. Stat.*, 27, 832-37, 1956.
- [125] Rosenberg J.V. and Schuermann T. , *A General Approach to Integrated Risk Management with Skewed, Fat-Tailed Risk*, Journal of Financial Economics, 2006.
- [126] Schlather M., *Models for stationary max-stable random fields. Extremes*, 5(1) : 33-44, 2002.
- [127] Schultze J. and Steinebach J., *On least squares estimates of an exponential tail coefficient, Statist. Decisions*, 14, 353-372, 1996.
- [128] Scott R. and Horvath P., *On the direction of preference for moments of higher order than the variance*, Journal of Finance 35, 915-919, 1980.
- [129] Sklar A., *Fonctions de répartition à n dimensions et leurs marges. Publications de L'Institut de Statistique de L'Université de Paris*, 8 :229-231,1959.
- [130] Smith R.L., *Extreme value analysis of environmental time series : an application to trend detection in ground-level ozone. Statistical Science*, 4(4) :367-393, 1989.
- [131] Smith R.L., *Estimating tails of probability lois. The Annals of Statistics* 3, p. 1174-1207, 1987.
- [132] Smith R.L., *Maximum likelihood estimation in a class of nonregular cases. Biometrika* 72, p. 67-90, 1985.
- [133] Stoica R., Chatelain F. and Sigelle M., *Inférence paramétrique pour les processus ponctuels marqués en analyse d'images. In Applications de la géométrie stochastique à l'analyse d'images. sous la direction de X. Descombes, Hermès* 2011.
- [134] Tawn J.A., *Bivariate Extreme Value Theorm Models and Estimation, Biometrika*, 75(3), p. 397-415, 1988.
- [135] Tawn J.A., *Modelling multivariate extreme value distributions. Biometrika*, 75(2), p. 245-253, 1990.
- [136] Thatcher R.A., *The Long Term Pattern of Adult Mortality and the Highest Attained Age*, Journal of the Royal Statistical Society Series A, 1999.
- [137] Thompson M.L., Reynolds J., Cox L.H., Guttorp P. and Sampson P.D., *A review of statistical methods for the meteorological adjustment of the tropospheric ozone. Atmospheric Environment*, 2001.

- [138] Tiago de Oliveira J., *Bivariate extremes : decisions*, Bull. Internat. Statist. Inst., vol.XLVI, Book 2, p. 241-251, 1975.
- [139] Tiago de Oliveira J., *Bivariate extremes : models and statistical decision*, Tech. Report no.14, Center for Stochastic Processes, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina, USA, 1982.
- [140] Tibiletti L., *A shortcut way of pricing default risk through zero-utility principle*, Journal of Risk and Insurance, vol. 73, n.2., 303-308, 2006.
- [141] Todorovic P. and Rousselle J., *Some problems on flood analysis*, *Water Resour. Res.*, 7, pp. 1144-1150, 1971.
- [142] Todorovic P. and Zelenhasic E., *A stochastig model for flood analysis* *Water Resour. Res.*, 6, pp. 1641-1648, 1970.
- [143] Thompson P., Cai Y., Reeve D. and Stander J., *Automated threshold selection methods for extreme wave analysis*, *Coastal Engineering*, 56, p. 1013-1021, 2009.
- [144] Tryon R.G. and Cruse T.A., *Probabilistic mseomechanics for high cycle fatigue life prediction. Journal of Engineering Materials and Technology-Transactions of the AMSE 122, 209-214*, 2000.
- [145] Von Mises R., *La distribution de la plus grande de n valeurs*, *Revue de Mathématique*, Union Interbalcanique, p. 141-160, 1936.
- [146] Von Mises R., *La distribution de la plus grande de n valeurs*, *American Mathematical Society*, RI, USA, vol. II, pp. 271-294, 1954.
- [147] Watson G.S., *Extreme values in samples from m-dependent stationary stochastic processes*, *Annals of Mathematical Statistics*, 25, p. 798-800, 1954.
- [148] Welsch R.E., *A Weak Convergence Theorem for Order Statistics from Strong-Mixing Processes*, *Ann. Math. Statist.*, *Annals of Mathematical Statistics*, 42, p. 1367-1646, 1971.
- [149] Weissman I., *Estimation of Parameters and Large Quantiles Based on the k-Largest Observations*, *J. Amer. Statist. Assoc.*, 73, 812-815, 1978.
- [150] Wilks S.S., *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *Ann. Math. Statist.* Volume 9, Number 1, p. 60-62, 1938.
- [151] Yi W. and Bier V. *An application of Copulas to Accident Precursor Analysis*, *Management Science*, 44 : S257-S270, 1998.

- [152] Xu B., *Co-mouvements extrêmes, interdépendance ou contagion ?*, Une analyse d'après la théorie des valeurs extrêmes, LAREefi, Université Montesquieu Bordeaux IV, 2008.
- [153] Zari T., *Contribution à l'étude du processus empirique de copule*, Thèse de doctorat de l'université paris 6, spécialité mathématique, 2010.
- [154] Zwiers F., Zhang X. and Feng Y., *Anthropogenic influence on long return period daily temperature extremes at regional scales*, *Journal of Climate*, Vol 24, p. 881-892 , 2011.
- [155] Zoglat A., El Adlouni S., Ezzahid E., Amar A., Okou C. G. and Badaoui F., *Statistical methods to expect extreme values : Application of POT approach to CAC40 return index*, *International Journal of Statistics and Economics*,10 (1), 1-13.
- [156] Zoglat A., El Adlouni S., Badaoui F., Okou C. G. and Amar A., *Managing hydrological risk with extreme modeling : application of peaks over threshold model to the Lekkous basin*. *Journal of Hydrologic Engineering*, Accepté janvier 2014.