

UNIVERSITE PARIS - DAUPHINE  
UFR MATHÉMATIQUES DE LA DÉCISION

MÉMOIRE DE SYNTHÈSE

en vue de l'obtention de l'

# HABILITATION À DIRIGER DES RECHERCHES

SPÉCIALITÉ : Mathématiques appliquées

présenté et soutenu publiquement par

Christian Y. ROBERT

le 23 Novembre 2009

Titre :

**QUELQUES CONTRIBUTIONS À LA STATISTIQUE  
DES EXTRÊMES ET DES PROCESSUS:  
APPLICATIONS EN ASSURANCE ET EN FINANCE**

---

## COORDINATEUR

M. Christian GOURIEROUX      Professeur, Université Paris-Dauphine

---

## RAPPORTEURS

M. Thomas MIKOSCH              Professeur, University of Copenhagen  
M. Philippe SOULIER              Professeur, Université Paris-Ouest Nanterre

---

## EXAMINATEURS

M. Marc HOFFMANN              Professeur, Université Marne-la-Vallée  
M. Jean-Paul LAURENT            Professeur, Université Lyon 1  
M. Christian P. ROBERT          Professeur, Université Paris-Dauphine  
M. Nizar TOUZI                    Professeur, Ecole Polytechnique

# Remerciements

Ce mémoire d'habilitation doit beaucoup à Christian Gourieroux. Il m'a toujours fait confiance et m'a beaucoup aidé dans mes choix professionnels. J'ai beaucoup appris en travaillant avec lui. Je lui exprime ma plus profonde gratitude pour toutes les heures qu'il m'a consacrées.

Je remercie les professeurs Thomas Mikosch et Philippe Soulier d'avoir accepté d'être rapporteurs de mon dossier d'habilitation. Je leur exprime toute ma reconnaissance pour l'intérêt et le temps qu'ils ont accordés à la lecture de mes travaux. Je tiens également à remercier les professeurs Marc Hoffmann, Jean-Paul Laurent, Christian P. Robert et Nizar Touzi qui me font l'honneur d'être membres du Jury. Depuis le moment où j'ai été élève à l'ENSAE, ils m'ont toujours aidé par leurs précieux conseils et leurs avis scientifiques, je leur en serai toujours reconnaissant.

Tous les travaux du dernier chapitre présentés dans ce mémoire ont été effectués en collaboration avec Mathieu Rosenbaum lorsqu'il partageait mon bureau à l'ENSAE. J'ai eu beaucoup de plaisir à travailler avec lui. Je le remercie de m'avoir emmené vers des contrées inconnues de la statistique des processus et pour toutes les passionnantes discussions que nous avons eues. Il est vrai que nous sommes souvent en désaccord, mais nous finissons toujours par trouver un terrain d'entente. J'ai eu aussi beaucoup de plaisir à travailler ou échanger avec Paul Doukhan, Chris Ferro, Laurence Lescourret, Johan Segers. Je les remercie pour toutes les séances de travail passées à réfléchir ensemble.

Je remercie également Alain Monfort, Sylviane Gastaldo et Pierre Biscourp pour m'avoir accueilli chaleureusement dans leurs établissements et pour m'avoir offert d'excellentes conditions de travail. Merci aussi à tous les permanents et à l'équipe administrative de l'ENSAE pour la bonne ambiance de travail qui régnait toujours à l'entresol.

Merci enfin à ma femme pour sa patience et ses encouragements.

# Table des matières

<b>Introduction</b>	<b>1</b>
Axes de recherche . . . . .	1
Organisation du mémoire . . . . .	1
Liste des travaux . . . . .	2
<b>1 Statistique des extrêmes : estimation de l'indice extrême et de la distribution de la taille des groupes d'extrêmes</b>	<b>4</b>
1.1 Problématique . . . . .	4
1.2 Estimation par bloc . . . . .	6
1.2.1 Estimation par bloc disjoint . . . . .	6
1.2.2 Estimation par bloc glissant de l'indice extrême . . . . .	8
1.3 Estimation par intervalle . . . . .	9
1.3.1 Distributions asymptotiques des durées entre dépassements de seuil . . . . .	9
1.3.2 Les estimateurs et leurs propriétés asymptotiques . . . . .	11
1.4 Indice extrême multivarié . . . . .	12
1.4.1 Caractérisations et propriétés de l'indice extrême multivarié . . . . .	13
1.4.2 Les estimateurs et leurs propriétés asymptotiques . . . . .	14
1.5 Conclusion et projets de recherche . . . . .	15
<b>2 Modélisation des extrêmes : étude des liens entre queue de distribution, renouvellement et mémoire longue</b>	<b>16</b>
2.1 Problématique et définitions . . . . .	16
2.2 Extrêmes et queue de distribution des processus ARCH(1) et $\alpha$ -ARCH(1) . . . . .	17
2.2.1 Extrêmes et queue de distribution du processus ARCH(1) . . . . .	17
2.2.2 Extrêmes et queue de distribution du processus $\alpha$ -ARCH(1) . . . . .	18
2.3 Extrêmes, queue de distribution et mémoire longue d'un processus à racine unitaire stochastique . . . . .	19
2.3.1 Définition et stationarité du processus . . . . .	19
2.3.2 Propriétés du processus croissance-effondrement . . . . .	21
2.4 Conclusion et projets de recherche . . . . .	24
<b>3 Événements extrêmes et assurance</b>	<b>25</b>
3.1 Problématiques et définitions . . . . .	25
3.2 Probabilité de ruine en présence de grands risques et avec versements de dividendes	26
3.2.1 Mécanisme de versement de dividendes . . . . .	26
3.2.2 Equivalent de la probabilité de ruine . . . . .	27
3.3 Dépendance extrême des pertes catastrophiques . . . . .	28

3.3.1	Le modèle à facteur . . . . .	29
3.3.2	Distributions bivariées des extrêmes . . . . .	30
3.4	Queues de distributions des modèles composés . . . . .	32
3.4.1	$N$ a une queue de distribution fine . . . . .	33
3.4.2	$N$ a une queue de distribution épaisse . . . . .	35
3.5	Conclusion et projets de recherche . . . . .	36
<b>4</b>	<b>Statistique des processus et séries financières haute fréquence</b>	<b>37</b>
4.1	Introduction . . . . .	37
4.1.1	Séries financières haute fréquence et bruit de microstructure . . . . .	37
4.1.2	Estimation de la volatilité intégrée en présence de bruit de microstructure . . . . .	39
4.1.3	Erreurs de couverture en présence de bruit de microstructure . . . . .	40
4.2	Le modèle avec zones d'incertitude . . . . .	41
4.2.1	Description du modèle . . . . .	41
4.2.2	Discussion . . . . .	43
4.3	Estimation de la volatilité et de la covolatilité intégrées . . . . .	44
4.3.1	Estimation de la volatilité intégrée . . . . .	44
4.3.2	Estimation de la covolatilité intégrée . . . . .	46
4.4	Erreurs de couverture en présence de bruit de microstructure . . . . .	47
4.4.1	Stratégie de couverture de référence . . . . .	47
4.4.2	Stratégies de couverture dans le modèle avec zones d'incertitude . . . . .	48
4.4.3	Distributions asymptotiques des erreurs de couverture . . . . .	48
4.5	Autres travaux et projets de recherche . . . . .	52
	<b>Bibliographie</b>	<b>53</b>

# Introduction

Ce mémoire présente de manière synthétique la majeure partie des travaux de recherche que j'ai pu mener depuis ma thèse. Dans cette introduction, je souhaite mettre en lumière les raisons qui ont conduit ma démarche et mes choix. J'ai toujours souhaité que mes travaux aient pour départ un problème concret issu du monde de la finance ou de l'assurance. Dans certains d'entre eux, j'ai choisi de développer des outils capables de mesurer des caractéristiques de phénomènes réels. Dans d'autres, j'ai choisi de proposer des modèles capables de reproduire un certain nombre de faits observés. Dans d'autres, enfin, j'ai essayé de participer à un effort de réflexion théorique engagé par d'autres auteurs pour comprendre et analyser des interactions complexes observées dans les séries financières. Le chemin pour aboutir à une connaissance approfondie et opérationnelle des risques financiers et d'assurance est encore long mais j'espère avoir apporté par mes travaux des idées qui aideront les personnes intéressées par ces problématiques à avancer un peu.

## Axes de recherche

Le titre de ma thèse était "Analyse des queues de distribution et des valeurs extrêmes en finance : applications aux séries financières haute-fréquence". Elle m'a fait goûter à la théorie des valeurs extrêmes et m'a confronté aux problèmes de la modélisation des séries financières haute fréquence. C'est donc tout naturellement que j'ai poursuivi dans ces deux axes de recherche.

Tout d'abord, j'ai développé mon intérêt pour la théorie des valeurs extrêmes dans trois directions. Je me suis intéressé à la statistique des valeurs extrêmes et plus particulièrement à la mesure du regroupement des valeurs extrêmes lorsqu'il existe de la dépendance entre les observations. Cela m'a conduit à chercher et à comprendre les liens qui pouvaient exister entre l'intensité des phénomènes extrêmes, leur regroupement sur des intervalles de temps courts et leur persistance sur des durées plus longues. Depuis que j'ai commencé à enseigner, j'ai effectué une grande majorité des mes enseignements en assurance. J'ai donc logiquement essayé d'appliquer aussi la théorie des valeurs extrêmes à des problématiques assurantielles où les sinistres de grande ampleur sont malheureusement trop nombreux pour la société, mais peut être profitables pour les compagnies d'assurance.

Plus récemment, j'ai décidé avec Mathieu Rosenbaum de m'intéresser à nouveau à la modélisation des séries financières haute fréquence. Nous avons choisi de changer de point de vue par rapport à la littérature existante qui avait recours aux processus ponctuels pour modéliser les données tick par tick. Nous avons préféré introduire un mécanisme de passage entre un prix efficient issu de la littérature des mathématiques financières et un prix observé pour ainsi tenter de réconcilier la théorie et la pratique.

## Organisation du mémoire

Les travaux présentés, dont le lecteur trouvera la liste plus bas, sont regroupés en 4 chapitres. Les chapitres peuvent être lus de manière indépendante.

Le premier chapitre se base sur cinq articles publiés. Il s'intéresse à l'estimation de l'indice extrême qui est une mesure de l'intensité du regroupement des extrêmes et de la distribution de la taille des groupes d'extrêmes. Nous avons cherché à proposer de nouveaux estimateurs plus simples à utiliser que ceux existants (même si les théories associées peuvent être dans certains cas complexes) tout en se fixant comme objectif de pouvoir toujours caractériser leur comportement asymptotique. Nous avons aussi étudié les distributions asymptotiques d'estimateurs basés sur les durées entre les extrêmes.

Le second chapitre se base sur trois articles publiés dans des revues d'économétrie ou de probabilités appliquées. Il a pour but de comprendre les liens qui peuvent exister entre dynamique d'un processus (Markovien), queue de la distribution stationnaire, regroupement des extrêmes, persistance et mémoire longue. Pour cela, nous étudions en particulier deux processus : le processus  $\alpha$ -ARCH(1) et un processus à racine unitaire stochastique. Dans le premier cas, nous montrons que la spécification "linéaire" de la variance conditionnelle est fondamentale pour obtenir des queues de distribution de type Pareto et un phénomène de regroupement des extrêmes. Dans le second cas, nous observons que la présence d'une queue de distribution épaisse est fortement reliée à un phénomène de persistance et de regroupement des extrêmes. Bien évidemment, les conclusions obtenues ne valent que pour les processus considérés.

Le troisième chapitre se base sur trois articles publiés dans des revues de probabilités appliquées, notamment à l'assurance. Nous y étudions trois problèmes concrets qui mettent en jeu des distributions à queue épaisse : Que devient le comportement asymptotique de la probabilité de ruine d'un assureur en présence de grands risques lorsque l'on autorise une politique de versement de dividendes? Quelles sont les structures "naturelles" de la dépendance entre branches d'assurance soumises à des événements catastrophiques? Quelles sont les queues de distribution des montants agrégés des sinistres lorsque le nombre de sinistres peut avoir une distribution à queue épaisse?

Le dernier chapitre se base sur quatre articles plus récents à paraître ou en révision ou en soumission dans des revues de mathématiques financières ou d'économétrie de la finance ou de statistique. Il s'intéresse à la modélisation des séries financières haute fréquence. Il propose un nouveau modèle appelé "Modèle avec zones d'incertitude" qui permet de reproduire de nombreux faits stylisés de ces séries et il propose d'étudier l'impact du bruit de microstructure sur l'estimation de la volatilité intégrée et sur une stratégie de couverture d'un produit dérivé.

## Liste des travaux

### Articles présentés

Chapitre 1: Statistique des extrêmes: estimation de l'indice extrême et de la distribution de la taille des groupes d'extrêmes

- A1. Robert, C. (2008). Estimating the multivariate extremal index function. *Bernoulli*, 14, 1027-1064.
- A2. Robert, C. (2009). Inference for the limiting cluster size distribution of extreme values. *The Annals of Statistics*, 37, 271-310.
- A3. Robert, C. (2009). Asymptotic distributions for the intervals estimators of the extremal index and the cluster-size distribution. *Journal of Statistical Planning and Inference*, 139, 3288-3309.
- A4. Robert, C., Segers, J. and Ferro, C., (2009). A sliding blocks estimator for the extremal index. *Electronic Journal of Statistics*, 3, 993-1020.
- A5. Robert, C. (2009). On asymptotic distribution of maxima of stationary sequences subject to random failure or censoring. *Statistics and Probability Letters*, 80, 134-142.

Chapitre 2: Modélisation des extrêmes: étude des liens entre queue de distribution, renouvellement et mémoire longue

- A6. Robert, C. (2000). Extremes of  $\alpha$ -ARCH models. *Lecture Notes in Statistics*, 147, 223-244.
- A7. Gouriéroux, C. and Robert, C. (2005). Stochastic unit root models. *Econometric Theory*, 26, 1052-1090.
- A8. Robert, C. (2007). Stochastic stability of some state-dependent growth-collapse processes. *Advances in Applied Probability*, 39, 1-32.

Chapitre 3: Evénements extrêmes et assurance

- A9. Robert, C. (2005). Asymptotic probabilities of an exceedance over renewal thresholds and an application to risk theory. *Journal of Applied Probability*, 42, 153-162.
- A10. Lescourret, L. and Robert, C. (2006). Extreme dependence of multivariate catastrophic losses. *Scandinavian Actuarial Journal*, 2006-4, 203-225.
- A11. Robert, C. and Segers, J. (2008). Tails of random sums of a heavy-tailed number of light-tailed terms. *Insurance: Mathematics and Economics*, 43, 85-92.

Chapitre 4: Statistique des processus et séries financières haute fréquence

- A12. Robert, C. and Rosenbaum, M. (2009). Volatility estimation under endogenous microstructure noise. A paraître dans *Mathematical Finance*.
- A13. Robert, C. and Rosenbaum, M. (2009). A new approach for the dynamics of ultra high frequency data: the model with uncertainty zones. En révision dans *Journal of Financial Econometrics*.
- A14. Robert, C. and Rosenbaum, M. (2009). On the microstructural hedging error. Soumis à *SIAM Journal on Financial Mathematics*.
- A15. Robert, C. and Rosenbaum, M. (2009). On the limiting spectral distribution of the covariance matrices of time-lagged processes. Soumis à *Journal of Multivariate Analysis*.

**Articles non présentés**

- A16. Robert, C. (1998). Mouvements extrêmes des séries financières haute fréquence. *Finance*, 19, 221-247.
- A17. Lescourret, L. and Robert, C. (2009). Transparency matters: Price formation in presence of order preferencing. En révision dans *Journal of Financial Markets*.

**Livre (non présenté)**

- L1. Denuit, M. and Robert, C. (2007). Actuariat des assurances de personnes: modélisation, tarification et provisionnement. Collection Audit-Actuariat-Assurance, *Economica*, Paris. 400 p.

# Chapitre 1

## Statistique des extrêmes : estimation de l'indice extrême et de la distribution de la taille des groupes d'extrêmes

### 1.1 Problématique

Un grand nombre de résultats en théorie des valeurs extrêmes peuvent être obtenus à l'aide des processus ponctuels. Par exemple, les distributions asymptotiques des statistiques d'ordre les plus élevées se déduisent très simplement de la loi limite du processus ponctuel de dépassements de seuil. Plus formellement, considérons une série strictement stationnaire de variables aléatoires,  $(X_n)_{n \geq 1}$ , de distribution stationnaire notée  $F$ . Nous supposons que, pour tout  $\tau > 0$ , il existe une suite de seuils  $(u_n(\tau))_{n \geq 0}$  telle que  $\lim_{n \rightarrow \infty} n\bar{F}(u_n(\tau)) = \tau$ , où  $\bar{F} = 1 - F$ . Il est nécessaire et suffisant pour l'existence d'une telle séquence que  $\lim_{x \rightarrow x_f} \bar{F}(x) / \bar{F}(x-) = 1$  où  $x_f = \sup \{u : F(u) < 1\}$  (voir le théorème 1.7.13 dans [LLR83]). Un choix naturel du seuil est donné par  $u_n(\tau) = F^{\leftarrow}(1 - \tau/n)$  où  $F^{\leftarrow}$  est l'inverse généralisée de  $F$ , c'est-à-dire  $F^{\leftarrow}(y) = \inf \{x \in \mathbb{R} : F(x) \geq y\}$ . Le processus ponctuel de dépassements de seuil  $N_n^{(\tau)}(\cdot)$  est défini par

$$N_n^{(\tau)}(B) = \sum_{i=1}^n 1_{\{i/n \in B, X_i > u_n(\tau)\}}$$

pour n'importe quel Borélien  $B \subset E := (0,1]$ .

Si  $(X_n)_{n \geq 1}$  est une suite de variables aléatoires indépendantes et équidistribuées, alors  $N_n^{(\tau)}$  converge faiblement vers un processus de Poisson homogène d'intensité  $\tau$  (voir par exemple le théorème 5.3.2 dans [EKM97]). Si l'hypothèse d'indépendance est relâchée, mais qu'une condition de dépendance de "longue portée" ( $\Delta(u_n(\tau))$  définie dans la section 3 de [A2]) est satisfaite, alors le processus ponctuel limite devient un processus de Poisson composé. Plus précisément, s'il existe une constante  $\theta \geq 0$  telle que

$$\lim_{n \rightarrow \infty} \mathbb{P}(N_n^{(\tau)}(E) = 0) = e^{-\theta\tau},$$

et s'il existe une distribution sur les entiers  $\pi$  telle que la distribution de  $N_n^{(\tau)}(B_n)$  avec  $B_n = (0, q_n/n]$  sachant qu'il y a au moins un excès parmi les  $q_n$  premières observations converge vers  $\pi$ , i.e.

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(N_n^{(\tau)}(B_n) = m \mid N_n^{(\tau)}(B_n) > 0\right) = \pi(m), \quad m \geq 1, \quad (1.1)$$



où  $(q_n)$  est une suite  $\Delta(u_n(\tau))$ -séparatrice (voir la section 3 de [A2]), alors le processus ponctuel limite est un processus de Poisson composé homogène d'intensité  $\theta\tau$  et de distribution de la taille des groupes d'extrêmes  $\pi$  (voir [HHL88]). La constante  $\theta$  est dénommée indice extrême et son inverse est égale à la moyenne de  $\pi$ . On peut montrer que  $\theta \leq 1$  et que le processus de Poisson composé devient un simple processus de Poisson lorsque  $\theta = 1$ .

Les articles [A2], [A3] et [A4] sont consacrés à l'estimation de l'indice extrême et de la distribution de la taille des groupes d'extrêmes. L'approche naturelle pour estimer  $\theta$  et  $\pi$  consiste à d'identifier des groupes d'extrêmes, et puis à évaluer pour chaque groupe d'extrêmes les caractéristiques d'intérêt à partir de ces groupes. Les deux méthodes les plus anciennes sont connues sous les termes de méthode par bloc et méthode par run. Le schéma de dégroupage par bloc consiste à choisir une taille de bloc  $r_n$  et à partitionner les  $n$  observations en  $k_n = \lfloor n/r_n \rfloor$  blocs ( $\lfloor x \rfloor$  désigne la partie entière de  $x$ ). Le schéma de dégroupage par run consiste à choisir une longueur de run,  $p_n$ , et à imposer que chaque paire d'excès qui sont séparés de moins de  $p_n$  observations appartiennent au même groupe d'extrêmes. La taille du bloc  $r_n$  et la longueur de run  $p_n$  sont appelées suites d'identification des schémas de dégroupage et jouent un rôle déterminant dans les propriétés asymptotiques des estimateurs. Il est important de noter que  $\theta$  et  $\pi$  ne dépendent pas de la constante  $\tau$  choisie.

Le problème de l'estimation de l'indice extrême a reçu une grande attention dans la littérature. Les premiers estimateurs par bloc et par run ont été construits en utilisant différentes caractérisations probabilistes de l'indice extrême (voir par exemple [EKM97] Section 8.1, [ANT00], [SW94]). Les estimateurs proposés sont caractérisés par deux suites: la suite des seuils  $u_n(\tau)$  et la suite d'identification des schémas de dégroupage. Leur principal défaut est leur dépendance très forte dans le seuil qui est fonction de la distribution stationnaire inconnue. L'estimation du seuil  $u_n(\tau)$  est compliquée puisque qu'il y a vraiment très peu d'observations qui peuvent a priori le dépasser (voir par exemple [Dr03]). Pour contourner cette difficulté, il a fallu considérer des seuils plus bas et des caractérisations de l'indice extrême adaptées (voir par exemple [Le83], [Ob87])

$$\theta = \lim_{n \rightarrow \infty} s_n \mathbb{P} \left( \max_{1 \leq i \leq r_n} X_i > u_{s_n}(\tau) \right) / (r_n \tau),$$

et

$$\theta = \lim_{n \rightarrow \infty} \mathbb{P} \left( \max_{2 \leq i \leq p_n} X_i \leq u_{s_n}(\tau) \mid X_1 > u_{s_n}(\tau) \right),$$

où  $s_n = o(n)$ ,  $r_n = o(s_n)$  et  $p_n = o(s_n)$ . Ces caractérisations ont motivé d'autres estimateurs [Hs91, Hs93a, WN98] dont le seuil  $u_{s_n}(\tau)$  peut être estimé par  $X_{\lceil n\tau/s_n \rceil:n}$  la  $\lceil n\tau/s_n \rceil$ -ième plus grande valeur parmi  $X_1, \dots, X_n$  ( $\lceil x \rceil$  désigne le plus petit entier plus grand que  $x$ ). Notons que ces estimateurs sont déterminés par deux suites également:  $r_n$  (ou  $p_n$ ) et  $s_n$ . Plus récemment, Ferro et Segers [FS03] (voir aussi [Fe03]) ont proposé une méthode basée sur les durées entre les dépassements de seuil qui permet d'éliminer les suites d'identification des schémas de dégroupage. Nous en reparlerons plus longuement dans la section 1.3. Ils ont étudié plusieurs estimateurs basés sur les moments des durées de dépassements de seuil.

Il y a par contre très peu d'articles qui ont étudié le problème de l'estimation de la distribution de la taille des groupes d'extrêmes. Dans [Hs91], la condition (1.1) est utilisée pour motiver le choix de l'estimateur par bloc suivant

$$\hat{\pi}_{n,1}(m; r_n, u_{s_n}(\tau)) = \frac{\sum_{j=1}^{k_n} \mathbf{1}\{Y_{n,j}(u_{s_n}(\tau)) = m\}}{\sum_{j=1}^{k_n} \mathbf{1}\{Y_{n,j}(u_{s_n}(\tau)) > 0\}}, \quad (1.2)$$

où  $Y_{n,j}(u_{s_n}(\tau)) = \sum_{i=(j-1)r_n+1}^{jr_n} \mathbf{1}\{X_i > u_{s_n}(\tau)\}$ ,  $s_n = o(n)$  et  $r_n = o(s_n)$ . Soit  $E_\pi(T) = \sum_{m=1}^{\infty} T(m) \pi(m)$  où  $T$  est une fonction définie sur  $\{1, 2, \dots\}$ . La convergence en probabilité des estimateurs de la forme suivante

$$\sum_{m=1}^{r_n} T(m) \hat{\pi}_{n,1}(m; r_n, X_{\lceil n\tau/s_n \rceil:n})$$

de  $E_\pi(T)$  est démontrée. Notons que ces estimateurs sont aussi déterminés par deux suites:  $r_n$  et  $s_n$ . Récemment, une nouvelle méthodologie a été proposée dans [Fe03]: un algorithme récursif permet de construire une suite d'estimateurs à partir des moments empiriques des distributions jointes des durées entre dépassements. Ces estimateurs ne dépendent que du choix d'une suite de seuil  $u_{r_n}(\tau)$ . Il est montré que ces estimateurs sont convergents pour des séries  $m$ -dépendantes.

## 1.2 Estimation par bloc

### 1.2.1 Estimation par bloc disjoint

Dans [A2], nous proposons des nouveaux estimateurs par bloc qui ont l'avantage de ne dépendre que de la taille des blocs  $r_n$ , le seuil étant déterminé à partir de  $r_n$  (et aussi, si souhaité, d'un paramètre laissé au libre choix de l'utilisateur). L'approche est la suivante. Tout d'abord nous estimons les probabilités associées au processus de Poisson composé. Dans un second temps, nous utilisons un algorithme de dégroupage basé sur l'algorithme de Panjer [Pa81] bien connu en actuariat. Cette idée a aussi été proposée récemment dans [BG03] et [BG04] lorsque les réalisations du processus sont parfaitement observées, ce qui n'est pas le cas ici.

Plus précisément, notons  $N_E^{(\tau)}$  la limite en loi de  $N_n^{(\tau)}(E)$  lorsque  $n \rightarrow \infty$  et  $p^{(\tau)} = (p^{(\tau)}(m))_{m \geq 0}$  sa distribution. Soit  $(\zeta_i)_{i \geq 1}$  une suite de variables aléatoires positives à valeurs entières, indépendantes et de même distribution que  $\pi$  et soit  $\eta(\theta\tau)$  une variable aléatoire de distribution de Poisson de paramètre  $\theta\tau$  et telle que  $\eta(\theta\tau)$  est indépendante de  $(\zeta_i)_{i \geq 1}$ . Nous avons alors  $N_E^{(\tau)} \stackrel{d}{=} \sum_{i=1}^{\eta(\theta\tau)} \zeta_i$ , avec la convention que la somme est égale à 0 si la réalisation de la variable  $\eta(\theta\tau)$  est égale à 0. La distribution de  $N_E^{(\tau)}$  est donnée par

$$p^{(\tau)}(0) = \mathbb{P}(\eta(\theta\tau) = 0) = e^{-\theta\tau}, \quad (1.3)$$

$$p^{(\tau)}(m) = \sum_{j=1}^m \mathbb{P}(\eta(\theta\tau) = j) \mathbb{P}\left(\sum_{i=1}^j \zeta_i = m\right) = \sum_{j=1}^m \frac{e^{-\theta\tau} (\theta\tau)^j}{j!} \pi^{*j}(m), \quad (1.4)$$

$m \geq 1$ , où  $\pi^{*j}$  est le  $j$ -ième produit de convolution de  $\pi$ , c'est-à-dire

$$\pi^{*j}(m) = \begin{cases} 0 & m < j, \\ \sum_{i_1 + \dots + i_j = m} \pi(i_1) \dots \pi(i_j) & m \geq j. \end{cases}$$

En théorie du risque, le montant agrégé des sinistres est souvent supposé avoir une distribution de Poisson composée. L'algorithme de Panjer [Pa81] est une méthode qui permet de calculer de manière récursive les probabilités associées à la distribution de Poisson composée à partir des probabilités associées à la distribution des sinistres qui est supposée discrète. Il est possible d'inverser l'algorithme et d'évaluer récursivement la probabilité  $\pi(m)$  à partir des  $p^{(\tau)}(j)$ ,  $j = 0, \dots, m$ , et des  $\pi(j)$ ,  $j = 0, \dots, m-1$ , de la manière suivante

$$\pi(m) = -\frac{(p^{(\tau)}(m) + m^{-1} \ln(p^{(\tau)}(0)) \sum_{j=1}^{m-1} j \pi(j) p^{(\tau)}(m-j))}{\ln(p^{(\tau)}(0)) p^{(\tau)}(0)}, \quad m \geq 1. \quad (1.5)$$

Expliquons maintenant plus précisément comment sont construits les estimateurs. Nous supposons par la suite que  $u_n(\tau) = F^{\leftarrow}(1 - \tau/n)$ . Nous divisons les observations  $\{1, \dots, n\}$  en  $k_n$  blocs de taille  $r_n$ ,  $I_j = \{(j-1)r_n + 1, \dots, jr_n\}$  pour  $j = 1, \dots, k_n$ , et un dernier bloc  $I_{k_n+1} = \{r_n k_n + 1, \dots, n\}$ . Le nombre d'observations dépassant le seuil  $u_{r_n}(\tau)$  à l'intérieur du bloc  $j$  est défini par

$$N_{r_n, j}^{(\tau)} = \sum_{i \in I_j} 1 \{X_i > u_{r_n}(\tau)\}, \quad j = 1, \dots, k_n.$$

Puisque, par hypothèse,  $\lim_{n \rightarrow \infty} \mathbb{E}[N_{r_n, j}^{(\tau)}] = \tau$ , le paramètre  $\tau$  peut être considéré comme le nombre moyen de dépassement de seuil par bloc de taille  $r_n$  au delà du seuil  $u_{r_n}(\tau)$ . La distribution empirique,  $p_n^{(\tau)}$ , du nombre de dépassements de seuil à l'intérieur d'un bloc est donnée par

$$p_n^{(\tau)}(m) = \frac{1}{k_n} \sum_{j=1}^{k_n} 1\{N_{r_n, j}^{(\tau)} = m\}, \quad m \geq 0.$$

Comme mentionné plus haut, la principale difficulté de cette approche est que le calcul de nombre d'excès nécessite la connaissance du seuil  $u_{r_n}(\tau)$  qui est basé sur la distribution stationnaire inconnue. On doit donc l'estimer à partir des données. On va choisir  $\hat{u}_{r_n}(\tau) = X_{[k_n \tau]:k_n r_n}$  et puis définir le nouvel estimateur de  $p^{(\tau)}$  par

$$\hat{p}_n^{(\tau)}(m) = \frac{1}{k_n} \sum_{j=1}^{k_n} 1\{\hat{N}_{r_n, j}^{(\tau)} = m\}, \quad m \geq 0,$$

où  $\hat{N}_{r_n, j}^{(\tau)} = \sum_{i \in I_j} 1\{X_i > \hat{u}_{r_n}(\tau)\}$ .

Les estimateurs des probabilités de la taille des groupes d'extrêmes peuvent maintenant être définis. Pour s'assurer que les estimateurs des probabilités sont positifs et que leur somme ne peut excéder 1, on va définir de manière récursive les estimateurs de la manière suivante

$$\hat{\pi}_n^{(\tau)}(m) = \max\left(0, \min\left(\chi_n^{(\tau)}(m), 1 - \sum_{j=1}^{m-1} \hat{\pi}_n^{(\tau)}(j)\right)\right), \quad m \geq 1,$$

où

$$\chi_n^{(\tau)}(m) = -\frac{\left(\hat{p}_n^{(\tau)}(m) + m^{-1} \ln(\hat{p}_n^{(\tau)}(0)) \sum_{j=1}^{m-1} j \hat{\pi}_n^{(\tau)}(j) \hat{p}_n^{(\tau)}(m-j)\right)}{\ln(\hat{p}_n^{(\tau)}(0)) \hat{p}_n^{(\tau)}(0)}.$$

Finalement, nous pouvons en déduire plusieurs estimateurs de l'indice extrême ou d'approximations de cette quantité. Ce paramètre apparaît en effet dans plusieurs conditions de moments associées aux distributions de  $N_E^{(\tau)}$  et  $\zeta_1$  (lorsque ces moments existent) :

$$\mathbb{P}(N_E^{(\tau)} = 0) = e^{-\theta\tau}, \quad \mathbb{E}[\zeta_1] = \theta^{-1}, \quad \mathbb{V}[N_E^{(\tau)}] = \theta\tau\mathbb{E}[\zeta_1^2].$$

Fixons un entier  $m \geq 1$ . Nous considérons les deux approximations suivantes de  $\theta$  :

$$\theta_2(m) = \frac{1}{\sum_{j=1}^m j \pi(j)} \quad \text{et} \quad \theta_3^{(\tau)}(m) = \frac{\sum_{j=0}^m (j-\tau)^2 p(j)}{\tau \sum_{j=1}^m j^2 \pi(j)}.$$

Des estimateurs de  $\theta$ ,  $\theta_2(m)$  et  $\theta_3^{(\tau)}(m)$  sont alors construits comme contre-parties empiriques :

$$\hat{\theta}_{1,n}^{(\tau)} = -\frac{\ln(\hat{p}_n^{(\tau)}(0))}{\tau}, \quad \hat{\theta}_{2,n}^{(\tau)}(m) = \frac{1}{\sum_{j=1}^m j \hat{\pi}_n^{(\tau)}(j)}, \quad \hat{\theta}_{3,n}^{(\tau)}(m) = \frac{\sum_{j=0}^m (j-\tau)^2 \hat{p}_n^{(\tau)}(j)}{\tau \sum_{j=1}^m j^2 \hat{\pi}_n^{(\tau)}(j)}.$$

Tous les estimateurs introduits ici sont déterminés par la suite  $r_n$  et le paramètre  $\tau$  qui peut être fixé arbitrairement à 1. Ils fournissent une alternative intéressante aux estimateurs introduits dans [FS03] et [Fe03] que nous décrirons dans la section 1.3 et qui nécessitent la donnée unique d'un seuil  $u = u_{r_n}(\tau)$ .

Sous la condition que  $r_n \rightarrow \infty$  et  $r_n = o(n)$  quand  $n \rightarrow \infty$  et sous des hypothèses énoncées dans [A2], nous montrons que ces estimateurs sont convergents et admettent une distribution limite qui n'est pas nécessairement Gaussienne en raison des contraintes imposées pour avoir des probabilités positives et une somme unitaire. Les preuves nécessitent une étude du processus empirique de queue sous des conditions de mélange adaptées aux extrêmes. Nous renvoyons à [A2] pour plus de détails.

### 1.2.2 Estimation par bloc glissant de l'indice extrême

La méthode par bloc proposée dans [A2] consiste à découper les observations en blocs disjoints et à calculer des statistiques pour chacun de ces blocs. Sur une idée proposée par Johan Segers, nous avons décidé de regarder le comportement de l'estimateur de l'indice extrême lorsque l'on considère désormais des statistiques calculées sur des blocs glissants. Nous avons alors obtenu une réduction simultanée du biais et de la variance asymptotiques.

Pour un entier  $r$ , posons

$$F_r(u) = \mathbb{P}(M_r \leq u), \quad \tau_r(u) = r\bar{F}(u) \quad \text{et} \quad \theta_r(u) = -\frac{\log F_r(u)}{\tau_r(u)}. \quad (1.6)$$

Par définition de l'indice extrême, nous avons donc que  $\theta = \lim_{r_n \rightarrow \infty} \theta_{r_n}(u_{r_n}(\tau))$ . Les estimateurs de  $\theta$  qui vont être proposés sont basés sur des analogues empiriques des fonctions  $F_r$  and  $\tau_r$ .

Pour des entiers  $0 \leq s < r$ , posons  $M_{s,r} = \max_{s < i \leq r} X_i$  et  $M_r = M_{0,r}$ . La fonction de distribution  $F_r$  du maximum  $M_r$  d'un bloc de taille  $r$  peut être estimé en utilisant les maxima calculés sur  $k = \lfloor n/r \rfloor$  blocs disjoints ou sur  $n - r + 1$  blocs glissants :

$$\hat{F}_{n,r}^{\text{dj}}(u) = \frac{1}{k} \sum_{i=1}^k I(M_{(i-1)r,ir} \leq u), \quad \hat{F}_{n,r}^{\text{sl}}(u) = \frac{1}{n-r+1} \sum_{i=0}^{n-r} I(M_{i,i+r} \leq u).$$

On peut se poser la question de savoir si utiliser des blocs glissants plutôt que des blocs disjoints conduit à une amélioration de l'estimateur. En effet, les  $n - r + 1$  blocs dans la définition de  $\hat{F}_{n,r}^{\text{sl}}(u)$  se recoupent et sont donc fortement dépendants, même dans le cas d'observations indépendantes et identiquement distribuées.

Définissons maintenant nos deux estimateurs de l'indice extrême. Posons

$$\hat{\tau}_{n,r}(u) = \frac{1}{k} \sum_{j=1}^k \sum_{i=1}^r I(X_{r(j-1)+i} > u) = \frac{1}{k} \sum_{i=1}^{rk} I(X_i > u), \quad (1.7)$$

les estimateurs sont définis de la manière suivante :

$$\hat{\theta}_{n,r}^{\text{dj}}(u) = -\frac{\ln \hat{F}_{n,r}^{\text{dj}}(u)}{\hat{\tau}_{n,r}(u)}, \quad \hat{\theta}_{n,r}^{\text{sl}}(u) = -\frac{\ln \hat{F}_{n,r}^{\text{sl}}(u)}{\hat{\tau}_{n,r}(u)}.$$

Il est important de noter que  $\hat{\theta}_{n,r_n}^{\text{dj}}(\hat{u}_{r_n}(\tau)) = \hat{\theta}_{1,n}^{(\tau)}$ . Dans un premier temps, nous nous intéressons à la loi asymptotique et la variance asymptotique de ces estimateurs. Nous avons le théorème suivant.

**Théorème 1.1.** *Sous des conditions précisées dans [A4], nous avons lorsque  $r_n \rightarrow \infty$  et  $r_n = o(n)$ ,*

$$\sqrt{k_n} \begin{pmatrix} \hat{\theta}_{n,r_n}^{\text{dj}}(u_{r_n}(\tau)) - \theta_r(u_{r_n}(\tau)) \\ \hat{\theta}_{n,r_n}^{\text{sl}}(u_{r_n}(\tau)) - \theta_r(u_{r_n}(\tau)) \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, \mathbf{V})$$

où  $\mathbf{V} = (v_{ij})_{i,j=1,2}^2$  est symétrique et

$$v_{11} = \frac{\theta^2}{\alpha^2} (e^\alpha - 1 - \alpha) + \frac{\theta^2 c^2}{\alpha}, \quad v_{22} = v_{12} = 2 \frac{\theta^2}{\alpha^3} \left( e^\alpha - 1 - \alpha - \frac{\alpha^2}{2} \right) + \frac{\theta^2 c^2}{\alpha},$$

avec  $\alpha = \theta\tau$ .

Il est important de noter que  $v_{22} \leq v_{11}$ , ce qui veut dire que l'estimateur par bloc glissant est plus efficient que sa version par bloc disjoint.

Nous nous intéressons maintenant aux biais de ces estimateurs. Si  $\hat{\theta}_{n,r}(u)$  dénote l'un des deux estimateurs, nous avons la décomposition suivante :

$$\mathbb{E}[\hat{\theta}_{n,r}(u)] - \theta = \left( \mathbb{E}[\hat{\theta}_{n,r}(u)] - \theta_r(u) \right) + (\theta_r(u) - \theta).$$

La composante  $\theta_r(u) - \theta$  est inhérente au processus  $(X_n)_{n \geq 1}$  lui-même. Dans [A4], nous donnons trois exemples où  $\theta_{r_n}(u_{r_n}(\tau)) - \theta = O(1/r_n)$  lorsque  $n \rightarrow \infty$ . La seconde composante  $\mathbb{E}[\hat{\theta}_{n,r}(u)] - \theta_r(u)$  peut être obtenue à l'aide d'un développement de Taylor au second ordre. Nous obtenons que, lorsque  $n \rightarrow \infty$ ,

$$k_n \left( \mathbb{E}[\hat{\theta}_{n,r_n}(u_{r_n}(\tau))] - \theta_{r_n}(u_{r_n}(\tau)) \right) \rightarrow \mu = \begin{cases} \mu_{\text{dj}} & := \theta(2\alpha)^{-1}(e^\alpha - 1) + \alpha^{-1}\theta c^2, \\ \mu_{\text{sl}} & := \theta\alpha^{-2}(e^\alpha - 1 - \alpha) + \alpha^{-1}\theta c^2, \end{cases}$$

Il faut remarquer que  $0 \leq \mu_{\text{sl}} \leq \mu_{\text{dj}}$ . Si de plus  $\theta_{r_n}(u_{r_n}) - \theta = o(1/k_n)$ , alors nous avons que, lorsque  $n \rightarrow \infty$ ,  $k_n(\mathbb{E}[\hat{\theta}_{n,r}(u)] - \theta) \rightarrow \mu$ .

Dans [A4], nous proposons un algorithme qui, pour un choix de taille de bloc fixé, donne un estimateur avec une variance minimale et un biais plus faible que l'estimateur de départ  $\hat{\theta}_{n,r_n}^{\text{sl}}$ . Par contre le choix de la taille des blocs  $r_n$  reste pour l'instant un problème ouvert.

## 1.3 Estimation par intervalle

Ferro et Segers [FS03] (voir aussi [Fe03]) ont proposé une méthode basée sur les durées entre dépassements de seuil qui permet d'éliminer les suites d'identification des schémas de dégroupage pour estimer l'indice extrême et la distribution de la taille des groupes d'extrêmes. Ils ont construit plusieurs estimateurs basés sur les moments des durées entre les extrêmes. Cependant, ils n'établissent la convergence (en probabilité) de leurs estimateurs que pour des séries  $m$ -dépendantes. Dans [A3], nous montrons la convergence sous des conditions beaucoup plus larges et donnons aussi les lois limites de leurs estimateurs.

### 1.3.1 Distributions asymptotiques des durées entre dépassements de seuil

#### Le processus de dépassements de seuil normalisé

Pour caractériser la distribution asymptotique des durées entre dépassements de seuil, nous étudions le processus ponctuel de dépassements de seuil normalisé  $N_{n,r_n}^{(\tau)}(\cdot)$  défini sur  $\tilde{E} := (0, \infty)$  par

$$N_{n,r_n}^{(\tau)}(B) = \sum_{i=1}^n 1_{\{i/r_n \in B, X_i > u_{r_n}(\tau)\}}, \quad (1.8)$$

pour n'importe quel Borélien  $B \subset \tilde{E}$ . Remarquons que la normalisation en temps par le facteur  $1/r_n$  et le choix du seuil  $u_{r_n}(\tau)$  ne sont pas standards. Ces choix permettent d'obtenir un processus ponctuel défini sur  $(0, \infty)$  et non plus sur  $(0, 1]$  comme dans [HHL88].

Posons  $C_K^+(\tilde{E})$  l'ensemble des fonctions continues, non négatives à support compact sur  $\tilde{E}$  et soit  $M_p(\tilde{E})$  l'espace des mesures ponctuelles définies sur  $\tilde{E}$ . Nous avons le théorème suivant.

**Théorème 1.2.** *Sous des conditions précisées dans [A3],  $N_{n,r_n}^{(\tau)}(\cdot)$  converge faiblement vers un processus de Poisson composé homogène  $N^{(\tau)}(\cdot)$  d'intensité  $\theta\tau$  et de distribution de la taille des groupes des extrêmes  $\pi$  dans  $M_p(\tilde{E})$ , i.e. vers un processus ponctuel de transformée de Laplace donnée par*

$$\mathbb{E} \left[ \exp \left( - \int_0^\infty f dN^{(\tau)} \right) \right] = \exp \left( - \theta\tau \int_0^\infty (1 - L(f(t))) dt \right)$$

où  $f \in C_K^+(\tilde{E})$  et  $L$  est la transformée de Laplace de  $\pi$ .

Une première conséquence (mais déjà bien connue) du théorème précédent est que les excès (les dépassements de seuil) apparaissent en groupes. Plus intéressant, ces groupes sont séparés par des variables aléatoires exponentielles de moyenne  $1/\theta\tau$  et les durées entre les extrêmes d'un même groupe sont asymptotiquement nulles. Il y a une proportion  $\theta$  de durées entre les extrêmes qui séparent les groupes et qui sont asymptotiquement positives (les durées inter-groupes) et une proportion  $(1 - \theta)$  de durées entre les extrêmes qui sont des durées entre extrêmes à l'intérieur des groupes et qui sont asymptotiquement nulles (les durées intra-groupes). Par conséquent, la distribution limite d'une durée entre deux extrêmes est asymptotiquement un mélange entre une distribution exponentielle de paramètre  $\theta\tau$  avec une probabilité  $\theta$  et un dirac en 0 avec une probabilité  $(1 - \theta)$  (voir aussi [FS03]).

Le nombre d'excès au delà du seuil  $u_{r_n}(\tau)$  parmi les observations  $X_1, \dots, X_n, N_{n,r_n}^{(\tau)}(\tilde{E})$ , est noté  $N_n$ .

### Distributions limites jointes des durées entre extrêmes

Soient  $1 \leq S_{1,n} < \dots < S_{N_n,n} \leq n$  les dates d'observation des dépassements de seuil (resp.  $S_{i,n}^* = S_{i,n}/r_n$  les dates normalisées) et soient  $T_{i,n} = S_{i+1,n} - S_{i,n}$ , pour  $i = 1, \dots, N_n - 1$ , les durées entre les dépassements de seuil (resp.  $T_{i,n}^* = T_{i,n}/r_n$  les durées normalisées). Remarquons que la normalisation par le facteur  $1/r_n$  diffère de la normalisation par le facteur  $\bar{F}(u_{r_n}(\tau))$  de Ferro et Segers qui est asymptotiquement équivalente à  $\tau/r_n$ .

**Corollaire 1.1.** *Sous des conditions précisées dans [A3], nous avons que, lorsque  $r_n \rightarrow \infty$  et  $r_n = o(n)$ , pour tous entiers  $m \geq 0$ ,  $1 \leq j_0, \dots, j_m$ ,*

$$(T_{j_0,n}^*, T_{j_0+j_1,n}^*, \dots, T_{j_0+j_1+\dots+j_m,n}^*) \xrightarrow{d} (T_{j_0}^*, T_{j_0+j_1}^*, \dots, T_{j_0+j_1+\dots+j_m}^*)$$

où

$$\begin{aligned} \mathbb{P}(T_{j_0}^* > s_{j_0}, T_{j_0+j_1}^* > s_{j_1}, \dots, T_{j_0+j_1+\dots+j_m}^* > s_{j_m}) \\ = \theta \left( \prod_{i=1}^m e_{j_i} \right) \exp(-\theta\tau \sum_{i=0}^m s_{j_i}), \quad \text{avec } s_{j_0} > 0, \dots, s_{j_m} > 0, \end{aligned} \quad (1.9)$$

et la suite  $(e_j)_{j \geq 1}$  est définie par la récursion suivante:  $e_1 = \pi(1)$  et

$$e_j = \pi(1)e_{j-1} + \dots + \pi(j-1)e_1 + \pi(j), \quad j \geq 2. \quad (1.10)$$

Nous en déduisons que, pour  $m = 0$  et  $j_0 = 1$ , nous avons  $\mathbb{P}(T_1^* = 0) = 1 - \theta$  et

$$\mathbb{P}(T_1^* > s) = \theta \exp(-\theta\tau s), \quad s > 0.$$

Pour  $m = 1$ ,  $j_0 = 1$  et  $j_j = j$ , nous avons  $\mathbb{P}(T_1^* = 0, T_{1+j}^* = 0) = 1 - \theta(2 - e_j)$  et

$$\mathbb{P}(T_1^* > s, T_{1+j}^* > t) = \theta e_j \exp(-\theta\tau(s+t)), \quad s > 0, t > 0.$$

En particulier, nous en déduisons que

$$e_j = \mathbb{P}(T_{1+j}^* > 0 | T_1^* > 0) = \sum_{k=1}^j \pi^{*k}(j),$$

c'est-à-dire, que  $e_j$  est la probabilité que la  $(1+j)$ -ième durée entre les dépassements de seuil soit une durée inter-groupe sachant que la première durée était aussi une durée inter-groupe.

De simples calculs donnent alors, pour  $j \geq 1$ ,

$$\begin{aligned}\mathbb{E}[T_1^*] &= 1/\tau, & \mathbb{E}[(T_1^*)^2] &= 2/(\theta\tau^2), \\ \mathbb{E}[T_1^*T_{1+j}^*] &= e_j/(\theta\tau^2), & \mathbb{E}[\min(T_1^*, T_{1+j}^*)] &= e_j/(2\tau).\end{aligned}\tag{1.11}$$

Par conséquent, les paramètres  $\theta$  et  $e_j$  peuvent s'exprimer de la manière suivante

$$\theta = \frac{2(\mathbb{E}[T_1^*])^2}{\mathbb{E}[(T_1^*)^2]}, \quad e_j = \frac{2\mathbb{E}[T_1^*T_{1+j}^*]}{\mathbb{E}[(T_1^*)^2]} = \frac{2\mathbb{E}[\min(T_1^*, T_{1+j}^*)]}{\mathbb{E}[T_1^*]}.\tag{1.12}$$

### 1.3.2 Les estimateurs et leurs propriétés asymptotiques

Les estimateurs de l'indice extrême et de la distribution de la taille des groupes d'extrêmes sont définis comme les contreparties empiriques des équations (1.10) et (1.12)

$$\begin{aligned}\hat{\theta}_n &= \frac{2((N_n - 1)^{-1} \sum_{i=1}^{N_n-1} T_{i,n}^*)^2}{(N_n - 1)^{-1} \sum_{i=1}^{N_n-1} (T_{i,n}^*)^2}, \\ \hat{\pi}_n(l) &= \hat{e}_{l,n} - \sum_{i=1}^{l-1} \hat{\pi}_n(i) \hat{e}_{l-i,n}, \quad l = 1, 2, \dots, \\ \tilde{\pi}_n(l) &= \tilde{e}_{l,n} - \sum_{i=1}^{l-1} \tilde{\pi}_n(i) \tilde{e}_{l-i,n}, \quad l = 1, 2, \dots,\end{aligned}$$

où l'on a utilisé la convention  $\sum_{i=1}^0 = 0$  et

$$\begin{aligned}\hat{e}_{l,n} &= 2 \frac{(N_n - 1 - l)^{-1} \sum_{i=1}^{N_n-1-l} T_{i,n}^* T_{i+l,n}^*}{(N_n - 1)^{-1} \sum_{i=1}^{N_n-1} (T_{i,n}^*)^2}, \quad l = 1, 2, \dots, \\ \tilde{e}_{l,n} &= 2 \frac{(N_n - 1 - l)^{-1} \sum_{i=1}^{N_n-1-l} \min(T_{i,n}^*, T_{i+l,n}^*)}{(N_n - 1)^{-1} \sum_{i=1}^{N_n-1} T_{i,n}^*}, \quad l = 1, 2, \dots\end{aligned}$$

Il est important de noter que ces estimateurs ne dépendent que du seuil choisi  $u (= u_{r_n}(\tau))$ . Ni la normalisation par le facteur  $1/r_n$ , ni la constante  $\tau$  ne modifient directement les estimateurs et donc a priori leur loi asymptotique. Enfin, puisque  $\theta \in ]0, 1[$  et que  $\pi$  est une distribution de probabilité, il est aussi important de considérer les estimateurs contraints suivants

$$\begin{aligned}\hat{\theta}_n^* &= 1 \wedge \hat{\theta}_n, \\ \hat{\pi}_n^*(l) &= \min \left\{ \max \left\{ 0, \hat{e}_{l,n} - \sum_{i=1}^{l-1} \hat{\pi}_n^*(i) \hat{e}_{l-i,n} \right\}, 1 - \sum_{i=1}^{l-1} \hat{\pi}_n^*(i) \right\}, \\ \tilde{\pi}_n^*(l) &= \min \left\{ \max \left\{ 0, \tilde{e}_{l,n} - \sum_{i=1}^{l-1} \tilde{\pi}_n^*(i) \tilde{e}_{l-i,n} \right\}, 1 - \sum_{i=1}^{l-1} \tilde{\pi}_n^*(i) \right\}.\end{aligned}$$

Notons que Ferro et Segers ont aussi introduit des légères modifications de ces estimateurs pour supprimer le biais du premier ordre, voir [FS03] et [Fe03].

Sous des hypothèses énoncées dans [A3], nous montrons que les estimateurs sont convergents et admettent une distribution limite qui n'est pas nécessairement Gaussienne à cause des contraintes imposées pour avoir des probabilités positives et de somme unitaire. Nous renvoyons à [A3] pour plus de détails. Donnons simplement un résultat pour les estimateurs de  $\theta$  et  $\pi(1)$  lorsque  $\pi(1) > 0$ .

**Théorème 1.3.** *Sous des conditions précisées dans [A3], nous avons lorsque  $r_n \rightarrow \infty$  et  $r_n = o(n)$ ,*

$$\begin{aligned}N_n^{1/2}(\hat{\theta}_n - \theta) &\xrightarrow{d} N(0, \theta^3 \sum_{j=1}^{\infty} j^2 \pi(j)) \\ N_n^{1/2}(\hat{\pi}_n(1) - \pi(1)) &\xrightarrow{d} N(0, 4\pi(1)\theta^{-1}) \\ N_n^{1/2}(\tilde{\pi}_n(1) - \pi(1)) &\xrightarrow{d} N(0, 2\pi(1)\theta^{-1}(1 - 2\pi(1)/3))\end{aligned}.$$

## 1.4 Indice extrêmeal multivarié

Dans [A1], nous proposons des estimateurs de l'indice extrêmeal multivarié et donnons leurs propriétés asymptotiques. L'indice extrêmeal multivarié relie la distribution asymptotique du vecteur des maxima des composantes d'une suite de variables aléatoires multivariées strictement stationnaires à celle obtenue dans le cas où les données sont indépendantes et ont, pour distribution, la distribution stationnaire. Il permet également d'avoir une idée du degré de regroupement des extrêmes de la série multivariée.

Plus précisément, soit  $(\mathbf{X}_n = (X_{n,1}, \dots, X_{n,d}))_{n \geq 1}$  une suite de variables aléatoires strictement stationnaires de fonction de distribution stationnaire  $F(\mathbf{x}) = \mathbb{P}(X_{n,i} \leq x_i, i = 1, \dots, d)$ ,  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ , et de distribution univariée marginale  $F_i(x) = \mathbb{P}(X_{n,i} \leq x)$ ,  $i = 1, \dots, d$ . Nous supposons qu'il existe une suite de fonctions de normalisation dans  $\mathbb{R}^d$  ( $\mathbf{u}_n(\tau) = (u_{n,1}(\tau_1), \dots, u_{n,d}(\tau_d))_{n \geq 1}$ ,  $\tau = (\tau_1, \dots, \tau_d) \in (0, \infty)^d$ ), telle que

$$\lim_{n \rightarrow \infty} n(1 - F_i(u_{n,i}(\tau))) = \tau, \quad \text{pour } \tau > 0, i = 1, \dots, d, \quad (1.13)$$

et, pour une fonction  $\tilde{H}$  non dégénérée :  $(0, \infty)^d \mapsto [0, 1]$ ,

$$\lim_{n \rightarrow \infty} n(1 - F(\mathbf{u}_n(\tau))) = -\ln \tilde{H}(\tau), \quad \text{pour } \tau \in (0, \infty)^d. \quad (1.14)$$

Cette condition d'existence ne suppose pas que  $F_i$  est dans un domaine d'attraction d'une loi des extrêmes car la normalisation ne doit pas être obligatoirement linéaire dans ce cas précis. Cependant la fonction  $\tilde{G}$  définie par  $\tilde{G}(\tau) = \tilde{H}(\tau_1^{-1}, \dots, \tau_d^{-1})$  pour  $\tau \in (0, \infty)^d$  doit nécessairement être une distribution multivariée des extrêmes que la normalisation soit linéaire ou non (voir [Pe97] Proposition 2.1). En particulier,  $\tilde{G}$  est une distribution continue avec des marginales ayant une distribution de Fréchet standard. Il est important de noter que  $-\ln \tilde{H}$  est une fonction homogène de degré 1, c'est-à-dire  $-\ln \tilde{H}(c\tau) = -c \ln \tilde{H}(\tau)$  pour tout  $c > 0$  et  $\tau \in (0, \infty)^d$ .

Soit maintenant  $M_{n,i} = \max(X_{1,i}, \dots, X_{n,i})$  le maximum de la  $i$ -ième composante et introduisons le vecteur des maxima par composante  $\mathbf{M}_n = (M_{n,1}, \dots, M_{n,d})$ . Si  $(\mathbf{X}_n)_{n \geq 1}$  est une suite de vecteurs aléatoires indépendants et équidistribués, alors (1.14) est équivalent à

$$\mathbb{P}(\mathbf{M}_n \leq \mathbf{u}_n(\tau)) = \mathbb{P}(M_{n,i} \leq u_{n,i}(\tau_i), i = 1, \dots, d) \rightarrow \tilde{H}(\tau), \quad \text{quand } n \rightarrow \infty.$$

Cette convergence peut être étendue aux séries strictement stationnaires en supposant une condition de longue portée  $D(\mathbf{u}_n(\tau))$  introduite dans [Hs89], qui est la version multivariée naturelle de la condition  $D(u_n(\tau))$ . Soit  $\mathbb{S}_+ \equiv \{\tau \in [0, \infty)^d : |\tau| = 1\}$  où  $|\tau|^2 = \sum_{i=1}^d \tau_i^2$ . Si  $D(\mathbf{u}_n(\tau))$  est satisfaite pour chaque  $\tau \in (0, \infty)^d$  et si  $P(\mathbf{M}_n \leq \mathbf{u}_n(\tau_0))$  converge quand  $n \rightarrow \infty$  pour chaque  $\tau_0$  dans l'intérieur de  $\mathbb{S}_+$ , alors il existe une fonction  $\theta : (0, \infty)^d \mapsto (0, 1]$  telle que (voir Proposition 2.3 dans [Pe97])

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{M}_n \leq \mathbf{u}_n(\tau)) = \tilde{H}(\tau)^{\theta(\tau)}, \quad \text{pour } \tau \in (0, \infty)^d. \quad (1.15)$$

Soit  $H(\tau) = \tilde{H}(\tau)^{\theta(\tau)}$ . La fonction  $G$  définie par  $G(\tau) = H(\tau_1^{-1}, \dots, \tau_d^{-1})$  est aussi une distribution multivariée des extrêmes et ses marginales ont une distribution de Fréchet standard. La fonction

$$\theta(\tau) = \frac{-\ln H(\tau)}{-\ln \tilde{H}(\tau)} \quad (1.16)$$

est l'indice extrêmeal multivarié de  $(\mathbf{X}_n)_{n \geq 1}$ .

L'estimation de cette fonction a reçu peu d'attention dans la littérature. Récemment, quelques estimateurs ponctuels ont été proposés mais leurs propriétés asymptotiques n'ont pas été étudiées. Par exemple, dans [SW96], Smith et Weissman introduisent la classe des processus  $M_4$  (Multivariate



Maxima of Moving Maxima) et montrent que l'indice extrêmeal multivarié de processus appartenant à une famille assez large peut être approché de manière très fine par l'indice extrêmeal multivarié d'un processus  $M_4$ . Cependant l'estimation de  $\theta$  à partir d'un processus  $M_4$  est compliquée car elle nécessite l'estimation d'un nombre infini de paramètres, sauf si des hypothèses complémentaires sont prises. Smith et Weissman donnent aussi une caractérisation clé de  $\theta(\tau)$  comme un indice extrêmeal d'une suite dépendante des composantes du vecteur normalisées pour avoir des marginales qui ont une distribution Fréchet standard (voir la proposition 2.1 dans [SW96] et la proposition 1.1 plus bas). Par conséquent, si l'on réussit à transformer les données de manière à ce que les marginales soient (asymptotiquement) des Fréchet standards, alors il est possible d'utiliser une méthode univariée pour estimer l'indice  $\theta(\tau)$ , c'est ce qui est proposé dans [Fe03] où une étude par simulation a été conduite pour étudier la précision de l'estimateur.

### 1.4.1 Caractérisations et propriétés de l'indice extrêmeal multivarié

L'indice extrêmeal multivarié défini à l'équation (1.16) peut aussi être caractérisé à l'aide d'un processus ponctuel de dépassement de seuils

$$N_n^{(\tau)}(B) = \sum_{l=1}^n 1 \{l/n \in B, \mathbf{X}_l \not\leq \mathbf{u}_n(\tau)\},$$

où  $B$  est un Borélien inclus dans  $E = (0,1]$  et  $\{\mathbf{X}_l \not\leq \mathbf{u}_n(\tau)\} = \bigcup_{i=1}^d \{X_{l,i} > u_{n,i}(\tau_i)\}$ . Contrairement au cas univarié, il y a plusieurs façons de définir des dépassements de seuils. Ici nous définissons un dépassement comme l'événement que l'une des composantes de  $\mathbf{X}_l$  dépasse son seuil respectif. Supposons que la condition (1.15) est vérifiée et que  $(\mathbf{X}_n)_{n \geq 1}$  satisfait la condition de longue portée  $\Delta(\mathbf{u}_n(\tau))$  introduite dans [Na94] (qui est une condition légèrement plus forte que la condition  $D(\mathbf{u}_n(\tau))$ ), alors une condition nécessaire et suffisante de convergence de  $N_n^{(\tau)}(\cdot)$  est la convergence de  $N_n^{(\tau)}((0; q_n/n])$  vers une distribution discrète,  $\pi^{(\tau)}$ , sachant qu'il y a au moins un dépassement de seuil, i.e.

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( N_n^{(\tau)}((0; q_n/n]) = k \mid N_n^{(\tau)}((0; q_n/n]) > 0 \right) = \pi^{(\tau)}(k),$$

où  $(q_n)_{n \geq 1}$  est une suite  $\Delta(\mathbf{u}_n(\tau))$ -séparatrice.  $\pi^{(\tau)}$  est la distribution de la taille des groupes d'extrêmes. Sous ces hypothèses, le processus ponctuel limite est un processus de Poisson composé homogène,  $N^{(\tau)}$ , d'intensité  $-\theta(\tau) \ln \tilde{H}(\tau)$  et de distribution de la taille des groupes d'extrêmes  $\pi^{(\tau)}$ . Il est important de remarquer que, sous des conditions supplémentaires peu restrictives, nous avons la caractérisation suivante de l'indice extrêmeal multivarié (voir [Hs89], [MF05a])

$$\theta(\tau) = \left( \sum_{k=1}^{\infty} k \pi^{(\tau)}(k) \right)^{-1},$$

i.e.  $\theta$  est égal à l'inverse du nombre moyen d'extrêmes par groupe.

Smith et Weissman ont aussi donné une caractérisation très intéressante de l'indice extrêmeal multivarié, voir [SW96]. Ils proposent de standardiser d'abord les marginales pour obtenir des Fréchet standards et puis ensuite d'exprimer  $\theta(\tau)$  comme l'indice extrêmeal univarié d'une suite qui est égale à une combinaison linéaire des composantes standardisées. Ici, nous préférons normaliser les marginales pour obtenir des Pareto standards.

**Proposition 1.1.** *Soit  $\tau \in [0, \infty)^d \setminus \{\mathbf{0}\}$  et supposons que  $\lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{M}_n \leq \mathbf{u}_n(\tau)) = \tilde{H}(\tau)^{\theta(\tau)}$ . Définissons la série univariée associée par*

$$Z_l^{(\tau)} = \max_{i=1, \dots, d} \tau_i Y_{l,i}, \quad l \geq 1,$$

où  $Y_{l,i} = (1 - F_{i,-}(X_{l,i}))^{-1}$  et  $F_{i,-}(x) = P(X_{l,i} < x)$ . Soit  $u_{Z^{(\tau)},n}(\varkappa) = \varkappa^{-1}(-\ln \tilde{H}(\tau))n$ . Alors  $\theta(\tau)$  est l'indice extrême univarié de la suite  $(Z_l^{(\tau)})_{l \geq 1}$ , i.e. il satisfait

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(Z_l^{(\tau)} > u_{Z^{(\tau)},n}(\varkappa)\right) = \varkappa \quad \text{et} \quad \lim_{n \rightarrow \infty} \mathbb{P}\left(\max_{l=1,\dots,n} Z_l^{(\tau)} \leq u_{Z^{(\tau)},n}(\varkappa)\right) = e^{-\theta(\tau)\varkappa}.$$

Il est bien connu que, dans le cas univarié, l'indice extrême ne dépend pas du choix de la constante  $\tau$ . Dans le cas multivarié,  $\theta(\tau)$  doit être constant sur toutes les demi-droites partant de l'origine. Il sera important de proposer des estimateurs satisfaisant cette propriété. Plus précisément, l'indice extrême multivarié a les propriétés suivantes (voir [Na94], [Pe97], la section 10.5.2 dans [BGST04], [MF05a] et [MF05b]):

1.  $0 \leq \theta(\tau) \leq 1$  pour tout  $\tau \in (0, \infty)^d$ .  $\theta(\tau)$  est une fonction continue de  $\tau \in (0, \infty)^d$  et est invariante par changement d'échelle, c'est-à-dire  $\theta(c\tau) = \theta(\tau)$  pour tout  $c > 0$  et  $\tau \in (0, \infty)^d$ .
2.  $\theta$  peut être étendu par continuité à  $[0, \infty)^d \setminus \{\mathbf{0}\}$ . Soit  $\tau^{(i)} = (0, \dots, 0, \tau_i, 0, \dots, 0)$ . L'indice extrême univarié,  $\theta_i$ , de la  $i$ -ième composante de  $(X_{n,i})_{n \geq 1}$  existe et  $\theta_i = \theta(\tau^{(i)})$ .  $\{\mathbf{0}\}$  est un point de discontinuité de  $\theta$  s'il existe  $i$  et  $j$  tels que  $\theta_i \neq \theta_j$ . Notons que les fonctions  $H$  et  $\tilde{H}$  peuvent être étendues par continuité à  $[0, \infty)^d$ . En particulier  $\tilde{H}(\tau^{(i)}) = e^{-\tau_i}$ ,  $H(\tau^{(i)}) = e^{-\theta_i \tau_i}$  et  $\tilde{H}(\mathbf{0}) = H(\mathbf{0}) = 1$ .
3. Des bornes de  $\theta(\tau)$  sont données par

$$\frac{\max(\theta_1 \tau_1, \dots, \theta_d \tau_d)}{-\ln \tilde{H}(\tau)} \leq \theta(\tau) \leq \frac{\theta_1 \tau_1 + \dots + \theta_d \tau_d}{-\ln \tilde{H}(\tau)}.$$

La borne haute correspond au cas où  $G$  a des composantes indépendantes et la borne basse au cas où  $G$  a des composantes totalement dépendantes.

## 1.4.2 Les estimateurs et leurs propriétés asymptotiques

Comme dans [A2], nous considérons un schéma par bloc et estimons les seuils pour chaque composante. Divisons les données  $\{1, \dots, n\}$  en  $k_n$  blocs de taille  $r_n$ ,  $I_j$  pour  $j = 1, \dots, k_n$ , et un dernier bloc  $I_{k_n+1}$ . Le nombre d'excès par bloc pour le  $j$ -ième bloc est défini par  $N_{r_n,j}^{(\tau)} = \sum_{l \in I_j} 1\{\mathbf{X}_l \not\leq \mathbf{u}_n(\tau)\}$  pour  $j = 1, \dots, k_n$ , où  $u_{r_n,i}(\tau) = F_i^{+}(1 - \tau/r_n)$ . Les seuils doivent en pratique être estimés sur les données. Comme dans [A2] nous considérons des estimateurs des seuils qui sont basés sur les statistiques d'ordre. Si  $0 < \tau \leq r_n$ , nous posons  $\hat{u}_{r_n,i}(\tau) = X_{[k_n \tau],i}$  où  $X_{k,i}$  est la  $k$ -ième plus grande valeur parmi  $X_{1,i}, \dots, X_{k_n r_n,i}$ . Si  $\tau = 0$ , nous posons  $\hat{u}_{r_n,i}(0) = \infty$ . Maintenant définissons  $\hat{N}_{r_n,j}^{(\tau)} = \sum_{l \in I_j} 1\{\mathbf{X}_l \not\leq \hat{\mathbf{u}}_n(\tau)\}$  pour  $\tau \in [0, r_n]^d$ .

Pour estimer l'indice extrême multivarié, il semble relativement naturel d'exploiter la caractérisation donnée par l'équation (1.16). Remarquons alors que, sous des conditions appropriées,

$$\lim_{n \rightarrow \infty} \mathbb{P}(N_n^{(\tau)}((0,1]) = 0) = \lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{M}_n \leq \mathbf{u}_n(\tau)) = H(\tau)$$

et que, par (1.14),

$$\lim_{n \rightarrow \infty} \mathbb{E}[N_n^{(\tau)}((0,1])] = \lim_{n \rightarrow \infty} n(1 - F(\mathbf{u}_n(\tau))) = -\ln \tilde{H}(\tau).$$

Utilisons maintenant les contreparties empiriques de  $H(\tau)$  et de  $-\ln \tilde{H}(\tau)$  définies par

$$\hat{H}_n(\tau) = \frac{1}{k_n} \sum_{j=1}^{k_n} 1_{\{\hat{N}_{r_n,j}^{(\tau)}=0\}} \quad \text{et} \quad -\ln \widehat{\tilde{H}}_n(\tau) = \frac{1}{k_n} \sum_{j=1}^{k_n} \hat{N}_{r_n,j}^{(\tau)} \quad \text{pour } \tau \in [0, r_n]^d.$$

On pourrait alors considérer  $-\ln \widehat{H}_n(\tau) / -\ln \widehat{\widehat{H}}_n(\tau)$  pour estimer  $\theta(\tau)$ . Mais cette fonction n'est pas invariante par changement d'échelle comme doit l'être l'indice extrêmeal multivarié. C'est pourquoi nous préférons introduire un premier estimateur qui satisfait la propriété d'homogénéité :

$$\widehat{\theta}_n^{(1)}(\tau) = \frac{-\ln \widehat{H}_n(\tau/K(\tau))}{-\ln \widehat{\widehat{H}}_n(\tau/K(\tau))}, \quad \tau/K(\tau) \in [0, r_n]^d \setminus \{\mathbf{0}\},$$

où  $K$  est une fonction de  $[0, \infty)^d \setminus \{\mathbf{0}\}$  vers  $(0, \infty)$  qui est homogène d'ordre 1. Par exemple, il est possible de considérer la famille  $K_{c,a}(\tau) = c(\sum_{i=1}^d |\tau_i|^a)^{1/a}$  pour  $a > 0$  et  $c > 0$ .

Le second estimateur est dérivé de la caractérisation proposée par Smith et Weissman. Considérons le nombre de dépassement de seuil de  $(Z_n^{(\tau)})_{n \geq 1}$  au delà du seuil  $u_{Z^{(\tau)}, n}(\varkappa)$

$$N_{Z^{(\tau)}, n}^{(\varkappa)} = \sum_{l=1}^n \mathbf{1}\{Z_l^{(\tau)} > u_{Z^{(\tau)}, n}(\varkappa)\}.$$

La proposition 1.1 implique que  $\lim_{n \rightarrow \infty} -\ln \mathbb{P}(N_{Z^{(\tau)}, n}^{(\varkappa)} = 0) = \theta(\tau) \varkappa$ . Nous pouvons alors réutiliser la méthodologie proposée dans [A2] pour construire un estimateur. Tout d'abord, nous remplaçons les  $Z_l^{(\tau)}$  par leur contreparties empiriques puisque les distributions marginales  $F_i$  sont inconnues. Soit  $R_{l,i}$  le rang de  $X_{l,i}$  parmi  $(X_{1,i}, \dots, X_{k_n r_n, i})$ . Dans le cas d'ex-æquo, le plus petit rang entre les ex-æquo sera utilisé. Nous définissons alors

$$\check{Z}_l^{(\tau)} = \max_{i=1, \dots, d} \tau_i \check{Y}_{l,i} \quad \text{où} \quad \check{Y}_{l,i} = \frac{k_n r_n}{k_n r_n + 1 - R_{l,i}}.$$

Nous introduisons ensuite le nombre d'excès de  $(\check{Z}_l^{(\tau)})_{l \in I_j}$  pour le bloc  $j$ :  $N_{\check{Z}^{(\tau)}, r_n, j}^{(\varkappa)} = \sum_{l \in I_j} \mathbf{1}\{\check{Z}_l^{(\tau)} > u_{\check{Z}^{(\tau)}, r_n}(\varkappa)\}$ . Comme précédemment,  $u_{\check{Z}^{(\tau)}, r_n}(\varkappa)$  est inconnu, mais il peut être estimé par  $\hat{u}_{\check{Z}^{(\tau)}, r_n}(\varkappa) = \check{Z}_{[k_n \varkappa]:k_n r_n}^{(\tau)}$  où  $\check{Z}_{[k_n \varkappa]:k_n r_n}^{(\tau)}$  est la  $[k_n \varkappa]$ -ième plus grande valeur parmi  $\check{Z}_1^{(\tau)}, \dots, \check{Z}_{k_n r_n}^{(\tau)}$ . Finalement, définissons  $\hat{N}_{\check{Z}^{(\tau)}, r_n, j}^{(\varkappa)}$  comme la contrepartie de  $N_{\check{Z}^{(\tau)}, r_n, j}^{(\varkappa)}$  où  $u_{\check{Z}^{(\tau)}, r_n}(\varkappa)$  est remplacé par  $\hat{u}_{\check{Z}^{(\tau)}, r_n}(\varkappa)$  et introduisons notre second estimateur

$$\widehat{\theta}_n^{(2)}(\tau) = -\varkappa^{-1} \ln \left( \frac{1}{k_n} \sum_{j=1}^{k_n} \mathbf{1}\{\hat{N}_{\check{Z}^{(\tau)}, r_n, j}^{(\varkappa)} = 0\} \right), \quad \tau \in [0, \infty)^d \setminus \{\mathbf{0}\}, \varkappa \in (0, r_n].$$

Il est important de noter que cet estimateur est invariant par changement d'échelle sans transformation sur  $\tau$ .

Sous des hypothèses énoncées dans [A1], nous montrons que ces estimateurs fonctionnels sont convergents et admettent pour limite des processus Gaussiens. Nous renvoyons à [A1] pour plus de détails.

## 1.5 Conclusion et projets de recherche

Dans [A5], nous étudions l'influence de la perte accidentelle de données ou de censure sur la distribution asymptotique du maximum. Nous montrons que dans certaines configurations où les probabilités de perte ou de censure sont fortement liées aux observations, les regroupements d'extrêmes peuvent disparaître et conduire à un indice extrêmeal unitaire. Il est donc important d'analyser précisément la qualité des données avant d'utiliser les estimateurs proposés dans les sections précédentes.

Nous souhaitons dans un avenir proche étudier un nouvel estimateur de l'indice extrêmeal basé sur une vraisemblance empirique des durées entre les dépassements afin d'obtenir un estimateur plus efficace. La difficulté consiste à identifier les durées inter-groupes et les durées intra-groupes.

## Chapitre 2

# Modélisation des extrêmes : étude des liens entre queue de distribution, renouvellement et mémoire longue

### 2.1 Problématique et définitions

Dans ce chapitre, nous présentons deux types de chaînes de Markov et étudions les liens qui existent entre les queues de la distribution stationnaire, la dynamique des extrêmes et la présence de mémoire longue ou de persistance.

Nous avons d'abord étudié dans [A6] le processus  $\alpha$ -ARCH(1) introduit dans [DG94] et dont un cas particulier est le processus ARCH(1). Nous montrons que la spécification "linéaire" de la variance conditionnelle est fondamentale pour obtenir des queues de distribution de type Pareto et un phénomène de regroupement des extrêmes.

Nous avons ensuite étudié dans [A7] et [A8] deux processus à racine unitaire stochastique. Nous présentons dans la suite uniquement le processus de [A8] car les résultats obtenus avec le processus de [A7] sont qualitativement du même type. Ce processus, bien que Markovien, présente des propriétés très proches de celles des modèles introduits dans [TL86] et [Pa99] et étudiés et comparés dans [HHS07]. Pour ces modèles, il existe une relation très forte entre l'indice de queue des distributions des durées entre les chocs et le paramètre de mémoire longue des processus. Nous montrons que notre chaîne de Markov peut être découpée naturellement en cycles indépendants et équidistribués et que les longueurs des cycles partagent les mêmes propriétés que les durées des chocs des deux processus introduits dans [TL86] et [Pa99]. La nouveauté des papiers [A7] et [A8] réside dans le fait que, contrairement à l'intuition, les propriétés de persistance et de mémoire longue peuvent être aussi partagées par une chaîne de Markov.

Avant de présenter nos résultats, nous introduisons préalablement quelques définitions et notations (voir aussi par exemple [MT93]). Notons  $\nu$  la mesure de Lebesgue et  $\mathcal{B}$  l'ensemble des Boréliens de  $\mathbb{R}^+$  (ou de  $\mathbb{R}$ ). Pour tout ensemble mesurable  $A \in \mathcal{B}$ , nous définissons le date d'entrée de la chaîne de Markov  $(X_n)_{n \geq 0}$  dans  $A$  par  $\tau_A = \inf \{n \geq 1 : X_n \in A\}$ , et le nombre de visites de  $A$  par  $\eta_A = \sum_{n=1}^{\infty} 1_{\{X_n \in A\}}$ .

- Une chaîne de Markov  $(X_n)_{n \geq 0}$  est  $\nu$ -irréductible si n'importe quel ensemble de mesure de Lebesgues strictement positive peut être atteint par la chaîne de Markov dans un temps fini en partant de n'importe quelle valeur. Cette condition se réécrit

$$\forall A \in \mathcal{B}, \nu(A) > 0 \implies L(x, A) = \mathbb{P}(\tau_A < \infty | X_0 = x) > 0, \quad \forall x \geq 0.$$

- Une chaîne de Markov  $(X_n)_{n \geq 0}$  est Harris-récurrente, si elle est  $\nu$ -irréductible et si pour

n'importe quel ensemble de mesure de Lebesgues strictement positive, la chaîne retourne dans cet ensemble un nombre infini de fois presque sûrement en partant de n'importe quelle valeur. Cette condition se réécrit

$$\forall A \in \mathcal{B}, \nu(A) > 0 \implies \mathbb{P}(\eta_A = \infty | X_0 = x) = 1, \quad \forall x \in A.$$

- Une chaîne de Markov Harris-récurrente a une unique mesure invariante. Si cette mesure est finie, on dit que la chaîne de Markov est positive Harris-récurrente. Sinon on parle de chaîne nulle-récurrente.

## 2.2 Extrêmes et queue de distribution des processus ARCH(1) et $\alpha$ -ARCH(1)

Nous commençons par rappeler les propriétés du processus ARCH(1) (voir par exemple la section 8.4 dans [EKM97]).

### 2.2.1 Extrêmes et queue de distribution du processus ARCH(1)

Nous considérons la classe des modèles ARCH(1) définis par la récurrence stochastique suivante :

$$X_n = \sqrt{\beta + \lambda X_{n-1}^2} Z_n, \quad n \geq 1$$

où les  $Z_n$  sont des variables aléatoires indépendantes et équidistribuées de loi Gaussienne standard. On pose

$$h(u) = \mathbb{E}[(\lambda Z^2)^u] = \frac{(2\lambda)^u}{\sqrt{\pi}} \Gamma\left(u + \frac{1}{2}\right), \quad u \geq 0.$$

La fonction  $h$  est strictement convexe et il existe une unique solution  $\kappa = \kappa(\lambda) > 0$  telle que  $h(\kappa) = 1$ . De plus

$$\kappa(\lambda) \begin{cases} > 1 & \text{si } \lambda \in (0,1), \\ = 1 & \text{si } \lambda = 1, \\ < 1 & \text{si } \lambda \in (1, 2e^\gamma), \end{cases}$$

où  $\gamma$  est la constante d'Euler et vaut approximativement 0,5772. Si  $\lambda$  est plus petit que  $2e^\gamma$ , il existe toujours une solution stationnaire.

**Proposition 2.1.** *Si  $\beta > 0$  et  $\lambda \in ]0, 2e^\gamma[$ ,  $(X_n)_{n \geq 0}$  est une chaîne de Markov positive Harris-récurrente.*

Les queues de la distribution stationnaire sont de type Pareto (voir par exemple [Go89] et [Ke73]).

**Proposition 2.2.** *Soit  $X$  une version stationnaire du processus. Nous avons*

$$\mathbb{P}(X > x) \underset{x \rightarrow \infty}{\sim} \frac{c}{2} x^{-2\kappa}$$

où

$$c = \frac{\mathbb{E} [((\beta + \lambda X^2)^\kappa - (\lambda X^2)^\kappa) Z^{2\kappa}]}{\kappa \mathbb{E} [(\lambda X^2)^\kappa \ln(\lambda Z^2)]}.$$

Intéressons nous maintenant à la dynamique des extrêmes et plus particulièrement au processus ponctuel de dépassement. Soit  $R(x) = \#\{j \geq 1 : \tilde{X} \prod_{i=1}^j (\lambda Z_i^2) > x\}$  où  $\mathbb{P}(\tilde{X} > x) = x^{-\kappa}$ ,  $x \geq 1$ , et définissons  $p_k = \mathbb{P}(R(1) = k)$ . Soit  $(v_n(\tau))_{n \geq 1}$  une suite de seuils telle que  $\lim_{n \rightarrow \infty} \mathbb{P}(X^2 > v_n(\tau)) = \tau$ . On définit le processus ponctuel de dépassement de  $X^2$  par

$$N_n^{(\tau)}(B) = \sum_{i=1}^n 1_{\{i/n \in B, X_i^2 > v_n(\tau)\}}.$$

Nous avons la proposition suivante (voir [HRRV89], [Pe94]).

**Proposition 2.3.** *Si  $\beta > 0$  et  $\lambda \in ]0, 2e^\gamma[$ , alors  $N_n^{(\tau)}$  converge vers un processus de Poisson composé homogène d'intensité  $\theta\tau$  et de distribution de la taille des groupes d'extrêmes  $\pi$  caractérisées par*

$$\theta = p_0 = \int_1^\infty \mathbb{P}\left(\bigvee_{j=1}^\infty \prod_{l=1}^j (\lambda Z_l^2) \leq x\right) \kappa x^{-\kappa-1} dx$$

et

$$\pi(k) = \frac{(p_{k-1} - p_k)}{p_0}, \quad k \geq 1.$$

Les queues de distribution et les extrêmes du processus AR(1) avec innovations ARCH(1) sont étudiés dans [BK01]. Les résultats sont qualitativement de même nature.

### 2.2.2 Extrêmes et queue de distribution du processus $\alpha$ -ARCH(1)

Nous considérons maintenant la classe des modèles  $\alpha$ -ARCH(1) définis par la récurrence stochastique suivante :

$$X_n = \sqrt{\beta + \lambda (X_{n-1}^2)^\alpha} Z_n,$$

où les  $Z_n$  sont des variables aléatoires indépendantes et équidistribuées de loi Gaussienne standard. Le cas  $\alpha = 1$  correspond au modèle ARCH(1) présenté plus haut. Si  $\alpha < 1$ , il existe toujours une solution stationnaire.

**Proposition 2.4.** *Si  $\beta > 0$ ,  $\lambda > 0$  et  $0 < \alpha < 1$ ,  $(X_n)_{n \geq 0}$  est une chaîne de Markov positive Harris-récurrente.*

Dans [A6], nous montrons que, si  $1/2 < \alpha < 1$ , les queues de la distribution stationnaire sont désormais de type Weibull.

**Proposition 2.5.** *Soit  $X$  une version stationnaire du processus. Supposons  $1/2 < \alpha < 1$ . Nous avons*

$$\mathbb{P}(X > x) \underset{x \rightarrow \infty}{\sim} \frac{d}{x^v} \exp\{-ex^\tau\}$$

où

$$\tau = 2(1 - \alpha), \quad e = \frac{\alpha^{-\frac{\alpha}{1-\alpha}}}{2\lambda(1-\alpha)}, \quad v = (1 - \alpha) \left[ \frac{\ln 2}{\ln \alpha} + 1 \right].$$

Dans [A6], nous avons aussi étudié le processus ponctuel de dépassement de  $X^2$  défini par

$$N_n^{(\tau)}(B) = \sum_{i=1}^n 1 \{i/n \in B, X_i^2 > v_n(\tau)\}$$

où  $(v_n(\tau))_{n \geq 1}$  est une suite de seuils telle que  $\lim_{n \rightarrow \infty} \mathbb{P}(X^2 > v_n(\tau)) = \tau$ . Nous obtenons le résultat suivant.

**Proposition 2.6.** *Si  $\beta > 0$ ,  $\lambda > 0$  et  $0 < \alpha < 1$ , alors  $N_n^{(\tau)}$  converge vers un processus de Poisson homogène d'intensité  $\tau$ .*

Par conséquent, la structure même de la variance conditionnelle conduit à des queues de distribution et un comportement des extrêmes très différents. La spécification ‘‘linéaire’’ est la seule qui conduise à des queues épaisses de type Pareto si l'innovation est Gaussienne. Dans ce cas là, les extrêmes ont tendance à se regrouper. Ceci laisse supposer qu'il existe des liens entre queues de distribution et agrégation des extrêmes comme cela a pu être montré dans le cas des processus ‘‘linéaires’’. Il faut noter cependant que, même dans le cas du processus ARCH(1), la persistance est faible puisque que les auto-corrélations décroissent nécessairement à une vitesse géométrique. Dans la section suivante, nous introduisons un processus Markovien où la persistance est beaucoup plus forte puisque l'on pourra obtenir une vitesse de décroissance hyperbolique.

## 2.3 Extrêmes, queue de distribution et mémoire longue d'un processus à racine unitaire stochastique

### 2.3.1 Définition et stationarité du processus

Nous considérons un processus de croissance-effondrement (CE)  $(X_n)_{n \geq 0}$  caractérisé par les propriétés suivantes :

- Les périodes de croissance: le processus croît comme une marche aléatoire avec des incréments positifs indépendants et équidistribués  $\varepsilon_n$ . Nous notons  $F$  la fonction de répartition de  $\varepsilon_1$ . Nous supposons que  $\varepsilon_1$  a une densité  $f$  strictement positive sur  $]0, \infty[$  et admet une transformée de Laplace définie par  $\Psi_{\varepsilon_1}(u) = \mathbb{E}[e^{-u\varepsilon_1}]$  et bornée sur un voisinage de 0. Cette condition implique que  $\varepsilon_1$  a une queue fine, au plus exponentielle. La moyenne de  $\varepsilon_1$  est notée  $\mu$ .
- Les époques d'effondrement: il existe une fonction  $\varphi$  telle que, à chaque date  $n$ , le processus s'effondre à 0 avec la probabilité  $1 - \varphi(X_{n-1})$ . Naturellement  $\varphi(0) = 1$ . Par souci de simplicité, nous supposons que  $\varphi$  est une fonction continue monotone de  $]0, \infty[$  vers  $]0, 1[$  telle que  $\inf_{x > 0} \varphi(x) = \underline{\varphi} > 0$ . Nous notons  $\bar{\varphi}$  le maximum de  $\varphi$  sur  $]0, \infty[$ .

Ainsi la dynamique du processus  $(X_n)_{n \geq 0}$  est déterminée par la récurrence stochastique suivante :

$$X_n = \begin{cases} X_{n-1} + \varepsilon_n & , \text{ avec la probabilité } \varphi(X_{n-1}), \\ 0 & , \text{ avec la probabilité } 1 - \varphi(X_{n-1}). \end{cases} \quad (2.1)$$

Le processus (CE)  $(X_n)_{n \geq 0}$  est une chaîne de Markov homogène de transition correspondant à un mélange entre un dirac et une distribution continue

$$\begin{aligned} \mathbb{P}(X_n = 0 | X_{n-1} = x_{n-1}) &= 1 - \varphi(x_{n-1}), & x_{n-1} &\geq 0, \\ \mathbb{P}(X_n \in ]x_n, x_n + dx_n | X_{n-1} = x_{n-1}) &= \varphi(x_{n-1})f(x_n - x_{n-1})dx_n, & x_n &\geq x_{n-1}. \end{aligned}$$

Par conséquent, la distribution stationnaire, si elle existe, a une masse ponctuelle  $\kappa$  en  $\{0\}$  et une fonction de densité  $l$  sur  $]0, \infty[$ . Ces quantités satisfont les équations suivantes

$$\kappa = \int_0^{+\infty} (1 - \varphi(x))l(x) dx, \quad l(x) = \kappa f(x) + \int_0^x \varphi(y)f(x-y)l(y) dy. \quad (2.2)$$

L'atome  $\{0\}$  est régénératif et peut être utilisé pour définir les dates de régénération suivantes :  $T_0 = 0$ ,  $T_n = \inf\{t > T_{n-1} : X_t = 0\}$ ,  $n \geq 1$ . De cette façon, la chaîne de Markov (CE) peut être décomposée en cycles indépendants et de même distribution  $(X_t)_{T_n \leq t < T_{n+1}}$ ,  $n \geq 1$ . Nous notons les longueurs des cycles  $D_n = T_{n+1} - T_n$ ,  $n \geq 1$ . Finalement, nous définissons la suite décroissante  $(p_n)_{n \geq 0}$  par  $p_0 = 1$  et

$$p_n = \mathbb{E}\left[\prod_{k=1}^n \varphi(S_k)\right], \quad n \geq 1,$$

où  $S_k = \varepsilon_1 + \dots + \varepsilon_k$ . Remarquons que, pour  $d \geq 1$ ,

$$\mathbb{P}(D_1 \geq d+1) = \mathbb{E}[\varphi(\varepsilon_{T_1+1})\dots\varphi(\varepsilon_{T_1+1} + \dots + \varepsilon_{T_1+d-1})] = \mathbb{E}\left[\prod_{k=1}^{d-1} \varphi(S_k)\right] = p_{d-1}.$$

Nous supposons dans un premier temps que  $\varepsilon_1$  a une distribution exponentielle de paramètre  $\alpha$  parce que des calculs analytiques peuvent être conduits dans ce cas. Nous verrons que les conditions obtenues pour ce cas particulier sont proches des conditions du cas général.

**Proposition 2.7.** *Supposons que  $f(x) = \alpha \exp(-\alpha x)$ . La distribution stationnaire existe et est unique si et seulement si*

$$\chi = \int_0^{+\infty} \exp\left(-\alpha \int_0^x (1 - \varphi(y)) dy\right) dx < \infty. \quad (2.3)$$

Si la condition (2.3) est satisfaite alors

$$l(x) = C_e \exp\left(-\alpha \int_0^x (1 - \varphi(y)) dy\right) \quad \text{et} \quad \kappa = 1 - \chi C_e,$$

où

$$C_e = \left( \int_0^{+\infty} (2 - \varphi(x)) \exp\left(-\alpha \int_0^x (1 - \varphi(y)) dy\right) dx \right)^{-1}.$$

Une distribution stationnaire existe si  $\varphi$  converge vers une constante strictement plus petite que 1 ou converge vers 1 suffisamment lentement. L'idée derrière cette condition est que la probabilité d'effondrement ne doit pas être trop petite de manière à ce que le processus puisse revenir vers 0 suffisamment souvent.

**Remarque 2.1.** Si  $\lim_{x \rightarrow \infty} (1 - \varphi(x))x = c$ , il existe (resp. il n'existe pas) une distribution stationnaire si  $c > \alpha^{-1} = \mu$  (resp.  $c < \mu$ ).

**Remarque 2.2.** Remarquons que  $N(\cdot) = \sum_{k=1}^{\infty} 1\{S_k \in (\cdot)\}$  est un processus de Poisson homogène d'intensité  $\alpha$ . D'après le théorème de Campbell,

$$\mathbb{E}\left[\exp\left(\int_{[0,x]} \ln \varphi dN\right)\right] = \mathbb{E}\left[\prod_{k=1}^{N([0,x])} \varphi(S_k)\right] = \exp\left(-\alpha \int_0^x (1 - \varphi(y)) dy\right).$$

Considérons le premier cycle  $(X_t)_{T_1 \leq t < T_2}$ . Il est facile de voir que la fonction de survie de la date d'effondrement de ce cycle,  $X_{T_2-1}$ , est donnée par

$$P(X_{T_2-1} > x) = \mathbb{E}\left[\prod_{k=1}^{N([0,x])} \varphi(S_{k+T_1} - S_{T_1})\right] = \mathbb{E}\left[\prod_{k=1}^{N([0,x])} \varphi(S_k)\right].$$

Par conséquent, la condition (2.3) est équivalente à la condition  $\chi = \mathbb{E}[X_{T_2-1}] < \infty$ . On peut également montrer que cette condition est aussi équivalente à la condition  $\mathbb{E}[D_1] < \infty$ .

Nous étudions maintenant la stabilité stochastique du processus dans le cas général.

**Proposition 2.8.** *i) La chaîne de Markov (CE) est  $\nu$ -irréductible.*

*ii) La chaîne de Markov (CE) est positive Harris-récurrente si  $\lim_{x \rightarrow \infty} (1 - \varphi(x))x = c \in (\mu, \infty]$ .*

*iii) La chaîne de Markov (CE) est nulle-récurrente si  $\lim_{x \rightarrow \infty} (1 - \varphi(x))x = c \in (0, \mu)$ .*

Nous voyons que la condition ii) de cette proposition est très proche de la condition obtenue dans le modèle avec des incréments de loi exponentielle (voir la remarque 2.1).

Une chaîne de Markov est fortement instable ou transiente si elle ne retourne qu'un nombre fini de fois dans un ensemble de taille "raisonnable" (voir [MT93] pour une définition mathématique). Notons que si  $\lim_{x \rightarrow \infty} (1 - \varphi(x))x = 0$ , notre chaîne de Markov peut être soit nulle-récurrente ou transiente.

Une condition nécessaire pour avoir une chaîne de Markov (CE) qui ne diverge pas à l'infini et retourne vers n'importe quel ensemble est que la longueur des cycles soit presque sûrement finie.

**Proposition 2.9.**  *$D_1$  est presque sûrement fini si  $1 - \varphi$  n'est pas intégrable en  $+\infty$ .*

Il est bien connu que cette condition n'est pas suffisante pour avoir un processus "stable" et l'on suppose généralement en plus que les cycles doivent avoir une moyenne finie (voir la remarque 2.2). Sous cette condition  $(X_n)_{n \geq 0}$  a une distribution limite (voir par exemple le théorème 1.2 dans [As87] p.170).

Par conséquent, des conditions suffisantes de stationnarité peuvent être obtenues à partir du comportement limite des probabilités  $p_n$ . Si  $\bar{\varphi} < 1$ , nous voyons immédiatement que  $p_n \leq \bar{\varphi}^n$  et que  $D_1$  a des moments de tous les ordres. Étudions plutôt le cas  $\bar{\varphi} = 1$ . Nous souhaiterions comparer



des conditions sur la convergence de la fonction  $\varphi$  vers 1 avec l'existence de  $\mathbb{E}[D_1] = 1 + \sum_{n=0}^{\infty} p_n$ . Pour cette raison, nous considérons uniquement les fonctions qui tendent vers 1 à une vitesse hyperbolique.

La proposition suivante donne des résultats sur le comportement des probabilités  $p_n$  lorsque  $1 - \varphi(x)$  est à variations régulières (voir la définition dans la première section du chapitre suivant).

**Proposition 2.10.** *Supposons que  $1 - \varphi(x)$  est à variations régulières d'indice  $\tau > 2/3$ . Il existe une constante  $C$  telle que*

$$p_n = C \exp\left(\sum_{k=1}^n \log \varphi(k\mu)\right) (1 + o(1)), \quad (2.4)$$

pour des grandes valeurs de  $n$ .

Nous déduisons de cette proposition qu'il existe une distribution stationnaire si

$$\sum_{n=0}^{\infty} \exp\left(\sum_{k=1}^n \log \varphi(k\mu)\right) < \infty.$$

Cette condition est équivalente à la condition (2.3) obtenue dans le modèle avec des incréments de loi exponentielle.

Le choix  $\tau > 2/3$  est purement technique. Lorsque l'on compare l'équation (2.4) avec la définition de  $p_n$ , nous voyons que, pour des grandes valeurs de  $n$ ,  $S_n$  peut être approximée par sa moyenne  $n\mu$ . Pour  $\tau > 2/3$ , la vitesse de convergence de  $\varphi$  vers 1 est suffisamment rapide pour que les fluctuations de  $S_n$  autour de  $n\mu$  soient négligeables. Ce n'est plus le cas si  $\tau \leq 2/3$ .

### 2.3.2 Propriétés du processus croissance-effondrement

#### Queue de la distribution stationnaire

Dans cette section, nous supposons que  $\lim_{x \rightarrow \infty} (1 - \varphi(x))x = c \in (\mu, \infty]$ , et nous notons  $X$  une variable aléatoire qui a la même distribution que la distribution stationnaire. De plus, nous supposons que la distribution de  $X_0$  est aussi la distribution stationnaire. Donnons une caractérisation des moments de cette distribution.

**Proposition 2.11.** *Soit  $g$  une fonction mesurable. Alors*

$$\mathbb{E}[g(X)] = \kappa \left( g(0) + \sum_{n=1}^{\infty} \mathbb{E} \left[ g(S_n) \prod_{k=0}^{n-1} \varphi(S_k) \right] \right),$$

où  $S_0 = 0$  et  $\kappa = \mathbb{P}(X = 0)$ .

Quand ils existent, les moments de la distribution stationnaire peuvent s'écrire comme des sommes infinies de moments spécifiques de la trajectoire de la marche aléatoire  $(S_n)_{n \geq 0}$ .

**Exemple 2.1.** *i) Si  $g = 1 - \varphi$ , nous obtenons*

$$\mathbb{E}[g(X)] = \kappa \sum_{n=1}^{\infty} \mathbb{E} \left[ (1 - \varphi(S_n)) \prod_{k=0}^{n-1} \varphi(S_k) \right] = \kappa \left( \sum_{n=1}^{\infty} (p_{n-1} - p_n) \right) = \kappa,$$

ce qui est équivalent à la première condition dans (2.2).

*ii) Si  $g = 1$ , nous obtenons*

$$\mathbb{E}[g(X)] = 1 = \kappa \left( 1 + \sum_{n=1}^{\infty} \mathbb{E} \left[ \prod_{k=0}^{n-1} \varphi(S_k) \right] \right) = \kappa \left( 1 + \sum_{n=1}^{\infty} p_{n-1} \right) = \kappa \mathbb{E}[D_1].$$

Nous déduisons de l'exemple 2.1 ii) que

$$\mathbb{E}[g(X)] = \frac{1}{\mathbb{E}[D_1]} \left( g(0) + \sum_{n=1}^{\infty} \mathbb{E} \left[ g(S_n) \prod_{k=0}^{n-1} \varphi(S_k) \right] \right) = \frac{1}{\mathbb{E}[D_1]} \mathbb{E} \left[ \sum_{t=T_1}^{T_2-1} g(X_t) \right].$$

Il s'agit de l'équation de stationnarité bien connue pour les processus régénératif (voir par exemple le théorème 1.2 dans [As87] p.170).

Regardons maintenant le comportement de queue de certaines distributions. Commençons par la distribution stationnaire.

**Proposition 2.12.** *i) Supposons que  $\varphi_l = \lim_{x \rightarrow \infty} \varphi(x) < 1$ . Soit  $(a, \infty)$  le plus grand ouvert (avec  $a < 0$ ) tel que*

$$\Psi_{\varepsilon_1}(u) < \infty, \quad \text{et } u \in (a, \infty).$$

*Si  $a$  est infini, supposons de plus que  $f \circ \ln$  est à variations rapides, c'est-à-dire,*

$$\lim_{x \rightarrow \infty} \frac{f(\ln(tx))}{f(\ln(x))} = 0 \quad \text{si } t > 1, \quad \lim_{x \rightarrow \infty} \frac{f(\ln(tx))}{f(\ln(x))} = \infty \quad \text{si } 0 < t < 1.$$

*Si  $a$  est fini, supposons de plus que  $f \circ \ln$  est à variations régulières d'indice  $a$  et que  $\lim_{\delta \searrow 0} \Psi_{\varepsilon_1}(a + \delta) = \infty$ . Soit  $\gamma$  l'unique constante positive telle que  $\Psi_{\varepsilon_1}(-\gamma) = (\varphi_l)^{-1}$ . Alors*

$$\mathbb{P}(X > x) = \psi(x) \exp(-\gamma x),$$

*où  $\psi \circ \ln$  est une fonction à variations lentes.*

*ii) Supposons que  $1 - \varphi(x) = x^{-\tau} L(x)$  où  $L$  est une fonction à variations lentes si  $2/3 < \tau < 1$  et  $L(x) = c(1 + o(1))$  si  $\tau = 1$ . Alors*

$$\mathbb{P}(X > x) = (1 + o(1)) \kappa \sum_{n=\lfloor x/\mu \rfloor}^{\infty} p_n,$$

*pour des grandes valeurs de  $x$ .*

Premièrement, notons que, si  $\varphi(x) = \varphi_c < 1$  est constant,  $l$  satisfait l'équation intégrale suivante :

$$l(x) = \kappa f(x) + \varphi_c \int_0^x f(y) l(x-y) dy.$$

Soit  $\gamma_c$  la constante positive telle que  $\Psi_{\varepsilon}(-\gamma_c) = (\varphi_c)^{-1}$  et définissons la mesure de probabilité  $G$  par  $G(dx) = \varphi_c \exp(\gamma_c x) f(x) dx$ . Nous avons

$$h(x) = \kappa \exp(\gamma_c x) f(x) + \int_0^x h(x-y) G(dy), \quad (2.5)$$

où  $h(x) = \exp(\gamma_c x) l(x)$ . L'équation (2.5) est une équation de renouvellement. D'après la théorie du renouvellement (voir par exemple [As87] p.155), nous en déduisons qu'il existe une constante positive  $D_c$  telle que  $\lim_{x \rightarrow \infty} h(x) = D_c$ . Nous avons ensuite

$$\mathbb{P}(X > x) = (1 + o(1)) \gamma_c^{-1} D_c \exp(-\gamma_c x),$$

pour des grandes valeurs de  $x$ . Par conséquent, lorsque  $\varphi$  est une fonction constante, la queue de la distribution stationnaire décroît à une vitesse exponentielle et  $\psi(x)$  converge vers une constante positive lorsque  $x$  tend vers l'infini. Mais ce n'est pas toujours le cas si  $\varphi$  n'est pas constante.

Deuxièmement, notons que dans le cas de la proposition 2.12 ii),  $X$  est presque sûrement fini si  $\mathbb{E}[D_1] < \infty$ , ce qui rebondit à nouveau sur le lien entre l'existence de la distribution stationnaire et l'existence d'un premier moment pour  $D_1$ . De plus, nous avons

$$\mathbb{P}(X > \mu k) = (1 + o(1)) \frac{\mathbb{E}[(D_1 - k)^+]}{\mathbb{E}[D_1]},$$

pour des grandes valeurs de  $k$ . C'est le résultat que nous aurions obtenu si  $\varepsilon_1 = \mu$  presque sûrement, ce qui signifie aussi que, dans ce cas, la vitesse de convergence de  $\varphi$  vers 1 est suffisamment rapide pour que les fluctuations de  $S_n$  autour de  $n\mu$  soient négligeables.

## Analyse de la persistance

La proposition suivante caractérise la structure de corrélation du processus stationnaire  $(\varphi(X_n))_{n \geq 0}$ .

**Proposition 2.13.** *La fonction d'autocovariances du processus  $(\varphi(X_n))_{n \geq 0}$  est donnée par*

$$\mathbb{Cov}(\varphi(X_n), \varphi(X_{n+k})) = \kappa^2 \sum_{l=k}^{\infty} p_l.$$

Par conséquent, dans le cas de la proposition 2.12 ii), la fonction d'autocovariances de  $(\varphi(X_n))_{n \geq 0}$  et la queue de la distribution stationnaire de  $(X_n)_{n \geq 0}$  ont le même comportement asymptotique

$$\lim_{k \rightarrow \infty} \frac{\mathbb{P}(X > \mu k)}{\mathbb{Cov}(\varphi(X_n), \varphi(X_{n+k}))} = \frac{1}{\kappa}.$$

Notons maintenant  $A_n = X_{T_{n+1}-1}$  la taille de l'effondrement du  $n$ -ième cycle et étudions la queue de distribution de cette variable. Evidemment, les  $A_n$ ,  $n \geq 1$ , sont des variables aléatoires positives indépendantes et équidistribuées.

**Proposition 2.14.** *i) Supposons que  $\varphi_l = \lim_{x \rightarrow \infty} \varphi(x) < 1$  et que  $\varphi$  est une fonction non-décroissante. Alors*

$$\mathbb{P}(A_1 > x) = (1 + o(1)) (1 - \varphi_l) \kappa^{-1} P(X > x),$$

pour des grandes valeurs de  $x$ .

*ii) Supposons que  $1 - \varphi(x) = x^{-\tau} L(x)$  où  $L$  est une fonction à variations lentes si  $2/3 < \tau < 1$  et  $L(x) = c(1 + o(1))$  si  $\tau = 1$ . Alors*

$$\mathbb{P}(A_1 > x) = (1 + o(1)) p_{\lfloor x/\mu \rfloor},$$

pour des grandes valeurs de  $x$ .

Remarquons que dans le cas de la proposition 2.14 ii), nous avons

$$\mathbb{P}(A_1 > \mu x) = (1 + o(1)) P(D_1 > x),$$

pour des grandes valeurs de  $x$ . Cela signifie que la distribution de la taille des effondrements a le même comportement de queue hyperbolique.

Supposons maintenant que  $\lim_{x \rightarrow \infty} (1 - \varphi(x)) x = c \in (3\mu, \infty]$  telle que le moment d'ordre 2 et les auto-corrélations du processus stationnaire existent.

La structure croissance-effondrement peut conduire à des phénomènes de corrélations persistentes appelés mémoire longue. Plus précisément, la définition de la mémoire longue que nous utilisons est la suivante.

**Definition 2.1.** *Un processus stationnaire  $(X_n)_{n \geq 0}$  est dit à mémoire longue si la somme des valeurs absolues de ses auto-corrélations*

$$\rho(k) = \frac{\mathbb{E}[X_n X_{n+k}] - \mathbb{E}[X_n] \mathbb{E}[X_{n+k}]}{\sqrt{\mathbb{V}(X_n)} \sqrt{\mathbb{V}(X_{n+k})}}$$

*diverge, c'est-à-dire  $\sum_{k=0}^{\infty} |\rho(k)| = \infty$ .*

Une autre définition souvent utilisée est que la fonction d'auto-corrélations soit une fonction à variations régulières d'indice plus petit que 1. Dans la proposition suivante, nous donnons un minorant de la somme des valeurs absolues des auto-corrélations.

**Proposition 2.15.** *Supposons que  $1 - \varphi(x) = x^{-\tau}L(x)$  où  $L$  est une fonction à variations lentes si  $2/3 < \tau < 1$  et  $L(x) = c(1 + o(1))$  si  $\tau = 1$ . Il existe une constante  $B$  telle que*

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n |\rho(k)| \geq B \mathbb{E} [D_1^4].$$

Nous en déduisons la condition pour avoir un processus à mémoire longue.

**Corollaire 2.1.** *Supposons que  $\lim_{x \rightarrow \infty} (1 - \varphi(x))x = c \in (3\mu, 4\mu)$ . Alors  $\sum_{k=0}^{\infty} |\rho(k)| = \infty$ .*

En conclusion, nous observons que, bien que le processus (CE) soit Markovien, il présente des propriétés très proches de celles des modèles introduits dans [TL86] et [Pa99]. La différence essentielle réside dans le fait que les durées des cycles sont endogènes et reliées très fortement aux valeurs du processus lui-même.

Notons finalement que des versions en temps continu du processus croissance-effondrement existent et ont été étudiées dans [EK03], [EK05] et [BPSZ06] et que la vitesse de convergence vers la distribution stationnaire du processus à racine unitaire stochastique introduit dans [A7] a été caractérisé dans [DFMS04].

## 2.4 Conclusion et projets de recherche

Les liens qui relient les queues de distribution, la dynamiques des extrêmes et la persistance des processus sont complexes et certainement spécifiques à chaque grande famille de processus. Néanmoins, comprendre les interactions entre ces phénomènes pourra ensuite aider à comprendre comment les propriétés des divers outils statistiques utilisés pour mesurer toutes ces grandeurs évoluent d'un processus à l'autre.

Dans [Li00], les processus du type de ceux introduits dans [TL86] et [Pa99] sont utilisés pour modéliser des volatilités d'actifs financiers et reproduire à la fois de la persistance et des changements de régimes. L'hypothèse d'un processus sous-jacent Markovien modifierait profondément les méthodologies de tarification des produits dérivés... Il pourrait aussi être intéressant de réfléchir à des procédures statistiques capables de discerner ce qui relève d'une mémoire longue générée par un processus fractionnaire ou par un processus du type précédent et, dans cette dernière hypothèse, de décider si un processus Markovien est adapté ou non.

## Chapitre 3

# Événements extrêmes et assurance

### 3.1 Problématiques et définitions

Dans ce chapitre consacré à l'assurance, nous nous intéressons à l'impact de phénomènes de grande ampleur (c'est-à-dire à des phénomènes capables d'engendrer des pertes financières très importantes) sur la solvabilité des compagnies, sur les dépendances extrêmes entre différentes branches et sur l'évaluation des pertes possibles lors d'une accumulation de sinistres. L'étude de ces phénomènes dans la littérature actuarielle n'est pas récente (on pourra consulter par exemple les ouvrages [Bu96], [EKM97] et [As00]) mais elle trouve aujourd'hui de nouvelles perspectives depuis l'augmentation du nombre de catastrophes naturelles observées ces dernières années et depuis la survenance de nouveaux risques à grande échelle que sont les attentats terroristes.

Dans ce chapitre, nous serons amenés à utiliser les définitions suivantes :

- Une fonction de distribution  $F$  sur  $]0, \infty[$  est dite à queue fine (*thin-tailed*, TT) si pour un  $\gamma > 0$

$$\int_0^{\infty} e^{\gamma x} dF(x) < \infty.$$

- Une fonction de distribution  $F$  sur  $]0, \infty[$  est dite à queue épaisse (*heavy-tailed*, HT) si pour tout  $\gamma > 0$

$$\int_0^{\infty} e^{\gamma x} dF(x) = \infty.$$

- Une fonction de distribution  $F$  sur  $]0, \infty[$  est dite à queue longue (*long-tailed*, LT) si

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x-y)}{\bar{F}(x)} = 1$$

pour tout  $y > 0$ , où  $\bar{F} = 1 - F$ .

- Une fonction de distribution  $F$  sur  $]0, \infty[$  est dite sous-exponentielle (*subexponential*,  $\mathcal{S}$ ) si

$$\lim_{x \rightarrow \infty} \frac{\bar{F}^{*2}(x)}{\bar{F}(x)} = 2.$$

- Une fonction de distribution  $F$  sur  $\mathbb{R}$  appartient à la classe  $\mathcal{S}^*$  si  $F(x) > 0$  pour tout  $x$  et

$$\lim_{x \rightarrow \infty} \frac{\int_0^x \bar{F}(x-y) \bar{F}(y) dy}{\bar{F}(x)} = \int_0^{\infty} \bar{F}(y) dy.$$

- Une fonction de distribution  $F$  sur  $]0, \infty[$  est dite à variations consistantes (*consistently varying tail*,  $\mathcal{C}$ ) si

$$\lim_{y \uparrow 1} \limsup_{x \rightarrow \infty} \frac{\bar{F}(xy)}{\bar{F}(x)} = 1.$$

- Une fonction de distribution  $F$  sur  $]0, \infty[$  est dite à variations régulières (*regularly varying tail*,  $\mathcal{R}$ ) s'il existe  $\alpha > 0$  tel que

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(xy)}{\bar{F}(x)} = y^{-\alpha}, \quad y > 0,$$

ou de manière équivalente tel que  $\bar{F}(x) = x^{-\alpha} L(x)$ ,  $x > 0$ , où  $L$  est une fonction à variations lentes (*slowly varying function*), c'est-à-dire une fonction satisfaisant

$$\lim_{x \rightarrow \infty} \frac{L(xy)}{L(x)} = 1, \quad y > 0.$$

## 3.2 Probabilité de ruine en présence de grands risques et avec versements de dividendes

### 3.2.1 Mécanisme de versement de dividendes

Nous considérons le modèle de risque suivant :

(i) Les sinistres,  $U_1, U_2, \dots$ , sont des variables aléatoires positives indépendantes et équidistribuées de moyenne  $\mu_U$ .

(ii) Les sinistres surviennent à des instants aléatoires  $s_1 < s_2 < \dots$  tels que les variables aléatoires positives  $T_n = s_n - s_{n-1}$  sont indépendantes et équidistribuées de moyenne  $\mu_T$ , et sont indépendantes de  $(U_n)_{n \geq 1}$ .

(iii) Le taux de prime est supposé être égal à un et l'équilibre tarifaire est assuré par la condition  $\mu_U < \mu_T$ .

Le processus de surplus (respectivement le processus des réserves) est défini par  $S^c$

$$S_t^c = \sum_{k=1}^{N_t} U_k - t \quad (\text{resp. } R_t = -S_t^c)$$

où  $N_t = \max\{n \geq 0 : s_n \leq t\}$  est le nombre de sinistres survenus avant la date  $t$ . On note  $u$  le niveau des fonds propres initiaux fournis par les actionnaires. Un problème classique de la théorie du risque consiste à trouver un équivalent de la probabilité de ruine sur un horizon infini

$$\psi(u) = \mathbb{P}\left(\max_{t \geq 0} S_t^c > u\right)$$

lorsque le niveau des fonds propres  $u$  tend vers l'infini (voir par exemple [As00], [EKM97]). Puisque la ruine ne peut avoir lieu (dans le cadre de cette idéalisation mathématique) que, lorsqu'un sinistre survient, la probabilité de ruine peut aussi se réécrire

$$\psi(u) = \mathbb{P}\left(\max_{n \geq 0} S_n > u\right)$$

où  $(S_n)_{n \geq 1}$  est une marche aléatoire d'incrément  $X_n = U_n - T_n$ . Nous notons  $F_X$  la fonction de répartition commune des  $X_n$ .

Une modification originale (mais raisonnable) de ce modèle est de considérer que des dividendes sont versés aux actionnaires lorsque le niveau des réserves est suffisamment haut. Les modèles à

versement de dividendes lorsque le niveau des réserves dépasse une barrière ont une histoire longue en théorie du risque (voir par exemple [Bu96]). Nous envisageons ici une situation de déclenchement de versement de dividendes différente.

Soit  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  une fonction positive telle que  $0 \leq \varphi(x) \leq x$ . La compagnie d'assurance considère un temps d'arrêt  $\sigma_1$  par rapport à la filtration  $(\mathcal{F}_n)_{n \geq 1}$  où  $\mathcal{F}_n = \sigma(U_1, T_1, \dots, U_n, T_n)$  et choisit de verser, lorsque le niveau des réserves  $-S_{\sigma_1}$  (juste après le paiement du sinistre) est suffisamment élevé (au moins positif), une part  $\varphi(-S_{\sigma_1})$  de ses réserves. L'autre part  $-S_{\sigma_1} - \varphi(-S_{\sigma_1})$  est conservée dans une marge de solvabilité et le niveau des réserves est remis (artificiellement) à 0. La même règle est alors utilisée pour définir une suite de temps d'arrêt et une suite de versements de dividendes  $\varphi(S_{\sigma_{k-1}} - S_{\sigma_k})$ .

Définissons alors le processus  $\{S_n^\varphi\}$  par

$$S_n^\varphi = S_n + \sum_{k \geq 1: \sigma_k < n} \varphi(S_{\sigma_{k-1}} - S_{\sigma_k}),$$

où  $\sigma_0 = 0$  et  $S_0 = 0$ . La probabilité de ruine devient

$$\psi(u) = \mathbb{P}(M^\varphi > u)$$

où  $M^\varphi = \max\{S_n^\varphi : n \geq 1\}$ .

L'objectif de [A9] est de fournir un équivalent de la probabilité de ruine  $\psi(u)$  lorsque la distribution des sinistres est à queue épaisse d'un type à préciser.

### 3.2.2 Equivalent de la probabilité de ruine

Nous commençons par donner un résultat intermédiaire. Soit  $(Y, N)$  et  $(Y_n, N_n)_{n \geq 1}$  des couples de variables aléatoires indépendants et identiquement distribués tels que  $N$  est une variable aléatoire positive de moyenne  $\nu$  et  $Y$  a une queue de distribution épaisse d'un type à préciser. On note  $F$  la distribution de  $Y$ . On considère la série  $(Z_n)_{n \geq 1}$  définie par  $Z_n = Y_n - \Sigma_{n-1}$ , où  $\Sigma_{n-1} = \sum_{k=1}^{n-1} N_k$  ( $\Sigma_0 = 0$ ), et on étudie la probabilité que le maximum  $M = \max_{n \geq 1} Z_n$  dépasse des grandes valeurs de  $x$ . Il est important de noter que cette probabilité est aussi la probabilité qu'il existe  $n$  tel que  $Y_n$  dépasse le seuil  $x + \Sigma_{n-1}$ , où  $(\Sigma_{n-1})_{n \geq 1}$  est une suite de seuils de renouvellement. Remarquons aussi que  $\Sigma_{n-1}$  est indépendante de  $Y_n$  mais peut être fonction de ces valeurs passées.

**Théorème 3.1.** *Supposons que  $\mathbb{E}[N^2] < \infty$  et que*

$$\lim_{x \rightarrow \infty} \mathbb{P}(Y > x) / \int_x^\infty \mathbb{P}(Y > u) du, \quad (3.1)$$

alors

$$\lim_{x \rightarrow \infty} \frac{\mathbb{P}(M > x)}{\int_x^\infty \bar{F}(u) du} = \nu^{-1}.$$

La condition (3.1) n'est pas une condition usuelle pour caractériser une distribution à queue épaisse. Cependant, on peut montrer que si  $F$  est  $LT$ , alors la condition (3.1) est satisfaite.

Si l'on pose  $Y_n = \max_{0 \leq k < \sigma_n - \sigma_{n-1}} (S_{k + \sigma_{n-1}} - S_{\sigma_{n-1}})$  et  $N_n = \rho(S_{\sigma_{n-1}} - S_{\sigma_n})$  où  $\rho(x) = x - \varphi(x)$ , alors il est facile de voir que

$$\mathbb{P}(M^\varphi > u) = \mathbb{P}(M > u).$$

D'après le théorème précédent, la probabilité de ruine possède l'équivalent suivant

$$\psi(u) \underset{u \rightarrow \infty}{\sim} (\mathbb{E}[\rho(-S_{\sigma_1})])^{-1} \int_u^\infty \mathbb{P}(Y > x) dx.$$

Il reste ensuite à trouver un équivalent de  $\int_u^\infty \mathbb{P}(Y > x) dx$  pour des grandes valeurs de  $u$ . L'article [FZ03] étudie le comportement des queues de distribution du maximum d'une marche aléatoire avec des incréments  $(X_n)_{n \geq 0}$  de type *LT* et de moyenne négative sur un intervalle de la forme  $[0, \sigma]$  où  $\sigma$  est un temps d'arrêt pour la filtration de la marche aléatoire. Les auteurs montrent que

$$\mathbb{P}(Y > x) \underset{x \rightarrow \infty}{\sim} \mathbb{E}[\sigma] \mathbb{P}(X > x),$$

où  $X$  a la même distribution que  $X_n$ . Nous pouvons alors en déduire le résultat suivant (voir pour les détails techniques [A9]).

**Théorème 3.2.** *Supposons que  $F_X \in \mathcal{S}^*$ , alors*

$$\lim_{x \rightarrow \infty} \frac{\psi(u)}{\int_x^\infty \bar{F}_X(u) du} = \frac{\mathbb{E}[\sigma]}{\mathbb{E}[\rho(-S_\sigma)]}.$$

Dans le cas où il n'y a pas de distribution de dividendes, la fonction  $\varphi$  est égale à 0, et, en utilisant l'identité de Wald, nous avons que  $\mathbb{E}[\rho(-S_{\sigma_1})] = -\mathbb{E}[S_{\sigma_1}] = -\mathbb{E}[\sigma_1]\mathbb{E}[X]$ , ce qui donne l'équivalent asymptotique usuel de la probabilité de ruine.

### 3.3 Dépendance extrême des pertes catastrophiques

Chaque année, les catastrophes naturelles causent des dommages financiers, économiques et humains très importants en France. Dans le passé, les assureurs n'ont jamais montré une réelle volonté à couvrir ces risques bien que la demande d'assurance ait augmenté régulièrement au cours du temps. L'une des raisons avancées est que l'évaluation précise des dommages potentiels est un exercice difficile, même si de nouvelles méthodes pour modéliser la fréquence et l'intensité des sinistres ont vu le jour récemment. Une des missions des actuaires, qui ont pour responsabilité la solvabilité des compagnies d'assurance, est d'évaluer précisément l'impact d'un événement majeur et de rechercher alors une couverture complémentaire en réassurance ou par titrisation.

Nous nous intéressons dans [A10] à un modèle où l'intensité de la catastrophe naturelle est un facteur commun aux pertes subies dans plusieurs branches d'une compagnie et nous examinons la dépendance extrême de ces pertes. Cette recherche a été motivée par l'observation des montants de sinistres agrégés de deux branches, dommages Automobile et dommages Multi-Risques habitation, occasionnés par 736 "tempêtes". Les deux valeurs extrêmes ont pour origine les tempêtes Lothar et Martin survenues en Décembre 1999 et qui ont causé les dommages les plus importants jamais enregistrés par l'assurance française.

Les données montrent un lien fort entre les montants des dommages des deux branches et suggèrent que l'intensité de la tempête induit une dépendance. Par exemple, lors de tels événements, les chutes d'arbres sont en effet des éléments responsables à la fois de dommages causés aux habitations ou aux véhicules (en particulier dans les agglomérations).

Comprendre les différences entre une analyse par branche et une analyse jointe est d'un intérêt certain pour comprendre les risques de faillite des compagnies et évaluer les pertes potentielles lors de scénarii de stress. Parmi les problèmes à résoudre, on peut considérer en particulier le problème de l'évaluation de la probabilité  $p$  associée à l'occurrence d'un événement dont la représentation conduirait à un point dans la zone hachurée de la figure 3.1, c'est-à-dire un événement qui conduirait à des dommages dans les deux branches plus grands que les deux seuils choisis. Comme aucun point n'est présent dans cette zone, il n'est pas possible d'utiliser la probabilité empirique. Dans le cas multidimensionnel, comme dans le cas unidimensionnel, si l'on doit évaluer des quantités dans les queues de distribution, le recours à la théorie des valeurs extrêmes semble judicieux et approprié.

A notre connaissance, [A10] est le premier papier à proposer un modèle à facteur pour construire des structures de dépendance des pertes multivariées et il introduit également les outils statistiques



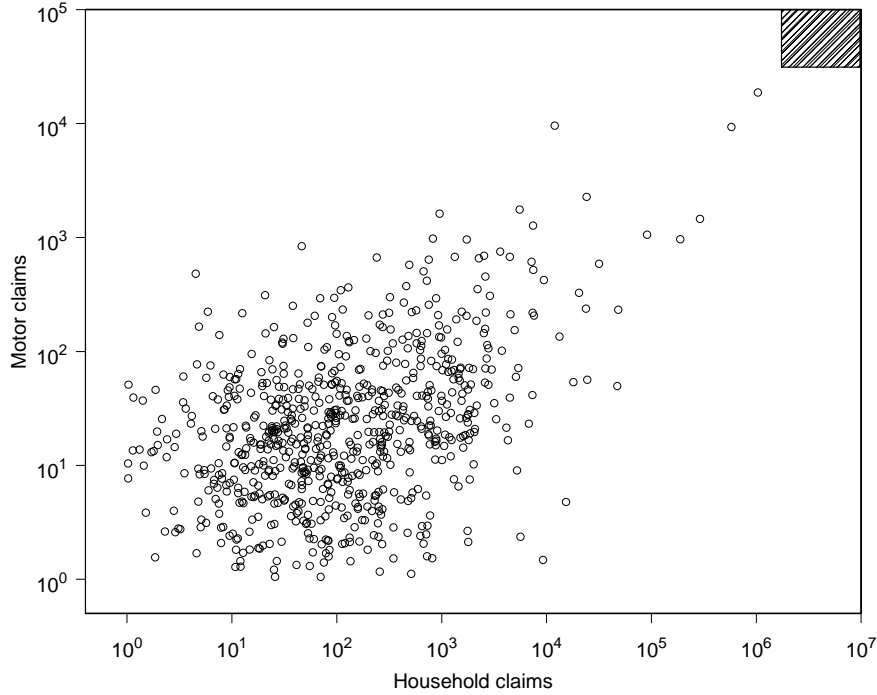


FIG. 3.1 – *Montants des sinistres dans les branches Automobile et Multi-risque habitation (échelle logarithmique)*

pour évaluer les probabilités de dépassement. Dans cette synthèse, nous nous bornerons à présenter le modèle et les structures de dépendance obtenues. Le lecteur est renvoyé à l'article pour des précisions sur les outils d'estimation et les applications aux données françaises.

### 3.3.1 Le modèle à facteur

Nous allons considérer un modèle où les distributions marginales des pertes sont des distributions de type Pareto car c'est ce que nous observons sur nos données. L'idée de base du modèle est de supposer l'existence d'un facteur commun relié à l'intensité du phénomène (voir aussi [LM03]). Nous notons  $X_{i,j}$  le montant des pertes de la  $j$ -ième branche pour la  $i$ -ième catastrophe naturelle. Nous considérons le modèle suivant :

$$X_{i,j} = T_j(Y_i \eta_{i,j}), \quad j = 1, 2, \quad (3.2)$$

où  $Y_i$  est l'intensité de la  $i$ -ième catastrophe naturelle, les  $\eta_{i,j}$  sont des variables aléatoires correspondant à des perturbations multiplicatives qui sont indépendantes de  $Y_i$ ,  $T_j$  est une fonction de transformation spécifique de la  $j$ -ième branche. Nous supposons que :

- $Y_i$  a une distribution de type Pareto  $F_Y$ , i.e. il existe une constante positive  $\alpha$  telle que

$$\bar{F}_Y(y) = 1 - F_Y(y) = y^{-\alpha} l_Y(y),$$

où  $l_Y$  est une fonction à variations lentes.

- $\eta_i = (\eta_{i,1}, \eta_{i,2})$  est un vecteur de variables aléatoires positives qui sont indépendantes de  $Y_i$ . De plus, il existe une constante  $\delta > \alpha$  telle que  $\mathbb{E}[\eta_{i,j}^\delta] < \infty$ ,  $j = 1, 2$ .

- $T_j$ ,  $j = 1, 2$ , sont des fonctions croissantes telles que leur inverse généralisée  $T_j^{\leftarrow}(x) = \inf \{y : T_j(y) \geq x\}$  sont des fonctions à variations régulières d'indice  $1/\gamma_j > 0$ , i.e.  $T_j^{\leftarrow}(x) = x^{1/\gamma_j} l_{T_j^{\leftarrow}}(x)$  avec  $l_{T_j^{\leftarrow}}$  une fonction à variations lentes.

-  $(Y_i, \eta_i)$  sont des vecteurs de variables aléatoires indépendants et identiquement distribués.

Cette construction implique que les montants des pertes  $X_{i,j}$ ,  $i \geq 1$ , sont des variables aléatoires indépendantes de distribution de type Pareto.

**Proposition 3.1.** *Supposons que  $X_{i,j}$  est spécifié par l'équation (3.2). Alors*

$$\bar{F}_{X_j}(x) = \mathbb{P}(X_{i,j} > x) = x^{-\beta_j} l_{X_j}(x), \quad i \geq 1, j = 1, 2,$$

où  $\beta_j = \alpha/\gamma_j$  et  $l_{X_j}$  est une fonction à variations lentes.

Dans notre modèle,  $Y_i$  induit de la dépendance entre les composantes du vecteur  $\mathbf{X}_i = (X_{i,1}, X_{i,2})$ . Concentrons-nous maintenant plus spécifiquement sur la dépendance extrême et donnons la distribution bivariée des extrêmes associée à ce modèle à facteur.

### 3.3.2 Distributions bivariées des extrêmes

#### Définitions

Soient  $(Y_{i,1}, Y_{i,2})$  pour  $i = 1, \dots, n$  des vecteurs aléatoires indépendants et équidistribués de distribution  $F$ . Pour  $j = 1, 2$  nous définissons  $M_{j,n} = \max(Y_{1,j}, \dots, Y_{n,j})$ . Les distributions bivariées des extrêmes apparaissent comme les limites du vecteur des maxima normalisés. Supposons qu'il existe des fonctions (linéaires)  $f_{1,n}, f_{2,n}$  telles que

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_{1,n} \leq f_{1,n}(y_1), M_{2,n} \leq f_{2,n}(y_2)) = G(y_1, y_2),$$

où  $G$  est une fonction de distribution non-dégénérée dans chacune de ses marginales.  $G$  est une distribution bivariée des extrêmes. Un choix approprié des fonctions de normalisation implique que les distributions marginales  $G_j$  doivent être l'une des trois distributions des extrêmes : la distribution de Weibull  $\Psi_\alpha(y) = \exp\{-(-\min(y, 0))^\alpha\}$ , pour  $\alpha > 0$ , la distribution de Gumbel  $\Lambda(y) = \exp\{-e^{-y}\}$ , ou la distribution de Fréchet  $\Phi_\alpha(y) = \exp\{-(\max(y, 0))^{-\alpha}\}$ , pour  $\alpha > 0$ .

Notons que, contrairement au cas univarié, il n'existe pas de familles paramétriques naturelles pour les distributions multivariées des extrêmes. Cependant  $G$  peut être caractérisée de plusieurs façons équivalentes. Par exemple, supposons que les distributions marginales de  $G$  sont des Fréchet standards  $\Phi_1(y) = \exp\{-y^{-1}\}$ ,  $y > 0$ . Une première représentation de  $G$  est donnée par

$$G(y_1, y_2) = \exp\left(-\int_{\mathcal{A}} \max(a_1 y_1^{-1}, a_2 y_2^{-1}) S(d\mathbf{a})\right), \quad (3.3)$$

où  $\mathbf{a} = (a_1, a_2)$  et  $S$  est une mesure finie non-négative sur  $\mathcal{A} = \{a_j \geq 0, j = 1, 2 : a_1 + a_2 = 1\}$  satisfaisant  $\int_{\mathcal{A}} a_j S(d\mathbf{a}) = 1$ ,  $j = 1, 2$  (voir par exemple la section 5.4 dans [Re87]). Une seconde représentation a été introduite par [Pi81] dans le cas bivarié :

$$G(y_1, y_2) = \exp\left(-\left(\frac{1}{y_1} + \frac{1}{y_2}\right) A\left(\frac{y_1}{y_1 + y_2}\right)\right), \quad (3.4)$$

où

$$A(w) = \int_0^1 \max(a(1-w), (1-a)w) S(da).$$

$A$  est une fonction convexe comprise entre  $\max(w, 1-w)$  et 1, et telle que  $A(0) = A(1) = 1$ . Elle est appelée fonction de dépendance. Voir aussi [Ga87], [Jo97], [HKK05] pour d'autres représentations.

## Distribution des extrêmes pour le modèle à facteur

Considérons maintenant notre modèle à facteur. Puisque les  $X_{i,j}$  ont des distributions de type Pareto, les distributions marginales de  $G$  sont des distributions de Fréchet. Il est bien connu que les choix des fonctions de normalisation  $f_{j,n}(x) = x^{1/\beta_j} U_{X_j}(n)$  où  $U_{X_j}(y) = \inf \{x : F_{X_j}(x) \geq 1 - y^{-1}\}$  conduisent à des distributions de Fréchet standards pour les distributions marginales de  $G$ , c'est-à-dire,

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_{j,n} \leq f_{j,n}(x_j)) = \Phi_1(x_j), \quad j = 1, 2.$$

(voir par exemple [EKM97]). Avant de donner la forme de la distribution limite, nous caractérisons la probabilité limite que les montants de pertes  $X_{i,j}$   $j = 1, 2$  dépassent les seuils  $f_{j,n}(x_j)$ . Ceci définit une notion de dépendance qui, n'est certes pas classique, mais est très utile dans la pratique. Pensons par exemple à un contrat de réassurance de type excédent de sinistre où le paiement du réassureur ne se mettrait en route que lorsque les deux montants dépasseraient les priorités.

**Proposition 3.2.** *Considérons le modèle (3.2). Pour  $x_1 > 0$  et  $x_2 > 0$ ,*

$$\lim_{n \rightarrow \infty} n\mathbb{P}(X_{i,1} > f_{1,n}(x_1), X_{i,2} > f_{2,n}(x_2)) = \Pi(x_1, x_2),$$

où

$$\Pi(x_1, x_2) = \mathbb{E} \left[ \min \left( x_1^{-1} \frac{\eta_{i,1}^\alpha}{\mathbb{E}\eta_{i,1}^\alpha}, x_2^{-1} \frac{\eta_{i,2}^\alpha}{\mathbb{E}\eta_{i,2}^\alpha} \right) \right]. \quad (3.5)$$

La distribution des extrêmes du modèle à facteur se déduit alors de la fonction  $\Pi$ .

**Corollaire 3.1.** *Considérons le modèle (3.2). Pour  $x_1 > 0$  et  $x_2 > 0$ ,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_{1,n} \leq f_{1,n}(x_1), M_{2,n} \leq f_{2,n}(x_2)) = \exp(-\Lambda(x_1, x_2)),$$

où

$$\Lambda(x_1, x_2) = \mathbb{E} \left[ \max \left( x_1^{-1} \frac{\eta_{i,1}^\alpha}{\mathbb{E}\eta_{i,1}^\alpha}, x_2^{-1} \frac{\eta_{i,2}^\alpha}{\mathbb{E}\eta_{i,2}^\alpha} \right) \right] = \frac{1}{x_1} + \frac{1}{x_2} - \Pi(x_1, x_2). \quad (3.6)$$

Puisque  $\Lambda(x_1, x_2) \neq x_1^{-1} + x_2^{-1}$ , les composantes du vecteur limite sont dépendantes. Cette dépendance joue un rôle très important dans les applications telles que la prédiction d'événements rares ou l'évaluation d'un événement conditionnellement à un autre. Il est à noter que les marginales de la distribution des extrêmes peuvent être indépendantes même si les composantes de la distribution de départ ne le sont pas (c'est le cas notamment de la distribution multivariée Gaussienne).

Remarquons que la dépendance extrême est engendrée par le vecteur  $\mathbf{Y}\eta_i = (Y_i\eta_{i,1}, Y_i\eta_{i,2})$ . La fonction de distribution  $F_{\mathbf{Y}\eta}$  est une fonction multivariée à variations régulières d'indice  $\alpha > 0$  (voir la section 5.4.2 dans [Re87]) et satisfait :

$$\lim_{t \rightarrow \infty} \frac{1 - F_{\mathbf{Y}\eta}(tx_1, tx_2)}{1 - F_{\mathbf{Y}\eta}(t, t)} = \frac{\mathbb{E} \left[ \max \left( x_1^{-\alpha} \eta_{i,1}^\alpha, x_2^{-\alpha} \eta_{i,2}^\alpha \right) \right]}{\mathbb{E} \left[ \max \left( \eta_{i,1}^\alpha, \eta_{i,2}^\alpha \right) \right]}.$$

Alors, en utilisant le corollaire 5.18 dans [Re87], il est possible de montrer qu'il existe une suite  $a_n$  telle que

$$\lim_{n \rightarrow \infty} \mathbb{P}(\max(Y_1\eta_{1,j}, \dots, Y_n\eta_{n,j}) \leq x_j a_n, j = 1, 2) = \frac{\mathbb{E} \left[ \max \left( x_1^{-\alpha} \eta_{i,1}^\alpha, x_2^{-\alpha} \eta_{i,2}^\alpha \right) \right]}{\mathbb{E} \left[ \max \left( \eta_{i,1}^\alpha, \eta_{i,2}^\alpha \right) \right]}.$$

Notre modèle est un petit peu plus général car il introduit une flexibilité par l'intermédiaire des fonctions de transformations  $T_j$  et permet d'avoir des indices  $\beta_j$  différents pour les marginales.

Remarquons finalement que  $\Pi$  et  $\Lambda$  sont deux fonctions homogènes d'ordre  $-1$ . Elles sont complètement déterminées par  $\mathcal{A}$  et sont équivalentes à la connaissance de  $A$  lorsque l'on veut caractériser la dépendance extrême.

Si  $\eta_{i,j}$  sont de plus des variables aléatoires indépendantes et équidistribuées, l'équation (3.6) définit une distribution des extrêmes bivariée symétrique. Notons  $H$  leur fonction de distribution et  $\eta$  une variable aléatoire générique de même distribution. Il est facile de voir que :

$$\begin{aligned}\Pi(x_1, x_2) &= x_1^{-1} \mathbb{E} \left[ \frac{\eta^\alpha}{\mathbb{E}[\eta^\alpha]} \bar{H} \left( \left( \frac{x_2}{x_1} \right)^{1/\alpha} \eta \right) \right] + x_2^{-1} \mathbb{E} \left[ \frac{\eta^\alpha}{\mathbb{E}[\eta^\alpha]} \bar{H} \left( \left( \frac{x_1}{x_2} \right)^{1/\alpha} \eta \right) \right] \\ \Lambda(x_1, x_2) &= x_1^{-1} \mathbb{E} \left[ \frac{\eta^\alpha}{\mathbb{E}[\eta^\alpha]} H \left( \left( \frac{x_2}{x_1} \right)^{1/\alpha} \eta \right) \right] + x_2^{-1} \mathbb{E} \left[ \frac{\eta^\alpha}{\mathbb{E}[\eta^\alpha]} H \left( \left( \frac{x_1}{x_2} \right)^{1/\alpha} \eta \right) \right],\end{aligned}$$

où  $\bar{H} = 1 - H$ . Ces deux caractérisations sont très utiles pour calculer analytiquement  $\Pi$  et  $\Lambda$ . Dans [A10], nous donnons des exemples de distributions des extrêmes bivariées symétriques bien connues et qui appartiennent à la famille définie par l'équation (3.6).

### 3.4 Queues de distributions des modèles composés

Dans le modèle collectif de la théorie du risque, le montant agrégé des sinistres d'un portefeuille,  $S_N$ , est modélisé par une somme de termes dont le nombre de termes est lui-même aléatoire :

$$S_N = \sum_{i=1}^N X_i$$

où  $N$  est une variable aléatoire à valeurs entières représentant le nombre de sinistres,  $X_1, X_2, \dots$  sont des variables aléatoires positives, indépendantes, équidistribuées et indépendantes de  $N$ , représentant les montants des sinistres.

L'étude des queues de distribution du montant agrégé des sinistres d'un portefeuille est très importante pour connaître le risque associé à ce portefeuille ou pour évaluer le coût associé à un programme de réassurance. Par exemple, si l'on considère un excédent de perte annuelle (Stop-loss) de rétention  $R$  qui rembourse  $(S_N - R)^+ = \max(0, S_N - R)$ , la prime pure associée est donnée par

$$\mathbb{E}[(S_N - R)^+] = \int_R^\infty \mathbb{P}(S_N > x) dx = \mathbb{P}(S_N > R) \mathbb{E}[S_N - R | S_N > R].$$

Elle peut se décomposer en deux parties : (i) la fréquence avec laquelle le traité est touché :  $\mathbb{P}(S_N > R)$ , (ii) le niveau de perte moyen lorsque le traité est touché :  $\mathbb{E}[S_N - R | S_N > R]$ . Pour des grandes valeurs de  $R$ , la fréquence de sévérité et la perte moyenne sont fortement liées au comportement de la queue de distribution de  $S_N$ .

En conditionnant par rapport au nombre de sinistres, nous obtenons

$$\mathbb{P}(S_N > x) = \sum_{n=1}^{\infty} \mathbb{P}(N = n) \mathbb{P}\left(\sum_{i=1}^n X_i > x\right),$$

ce qui montre que la queue de distribution de  $S_N$  dépend de la queue de distribution du nombre de sinistres  $N$  et aussi de celle des coûts des sinistres individuels. Une partie de la littérature actuarielle s'est penchée sur l'étude des queues de ces distributions. Les résultats dépendent fortement de l'hypothèse de queue de distribution fine ou de queue de distribution épaisse qui porte sur  $N$  et/ou sur  $X_1$ . On peut essayer de résumer très grossièrement les résultats dans le tableau suivant

$$\mathbb{P}(S_N > x) \underset{x \rightarrow \infty}{\sim}$$

$N \setminus X$	Queue fine	Queue épaisse
Queue fine	forme mixte	$\mathbb{E}[N] \mathbb{P}(X_1 > x)$
Queue épaisse	$\mathbb{P}(\mathbb{E}[X_1] N > x)$ ou ?	

TAB. 3.1

Nous allons présenter dans la suite chaque cas succinctement

### 3.4.1 $N$ a une queue de distribution fine

Deux cas sont à étudier séparément:  $X_1$  a une queue de distribution fine ou  $X_1$  a une queue de distribution épaisse.

#### $N$ a une queue de distribution fine et $X_1$ a une queue de distribution fine

Les résultats sont assez anciens et ont été fournis de manière disparate. Nous proposons de présenter ici une manière unifiée permettant de retrouver la très grande majorité des résultats établis dans la littérature (voir par exemple [BPP84], [Te85], [Su85], [EJMT85], [Je95], [CG00]). Pour une variable aléatoire positive  $\xi$ , nous définissons le domaine d'existence de sa fonction génératrice des moments  $\varphi_\xi(\lambda) = \mathbb{E}[e^{\lambda\xi}]$  par  $\Lambda_\xi := \{\lambda \in \mathbb{R} \mid \varphi_\xi(\lambda) < \infty\}$ . Notons que la fonction génératrice des moments de  $S_N$  est donnée par

$$\varphi_{S_N}(\lambda) = \varphi_N(\ln \varphi_{X_1}(\lambda)).$$

Nous supposons que  $\Lambda_N, \Lambda_{X_1}, \Lambda_{S_N}$  sont des ouverts qui contiennent 0. Des exemples sont donnés dans le tableau suivant:

Distribution de $N$		$\mathbb{P}(N = n)$	$\varphi_{S_N}(\lambda)$
Binomiale $(p, m)$	$p \in ]0, 1[$ $m \in \mathbb{N}$	$C_m^n p^n (1-p)^{m-n}$	$(1-p + p\varphi_{X_1}(\lambda))^m$
Poisson $(\mu)$	$\mu > 0$	$e^{-\mu} \mu^n / n!$	$e^{\{\mu(\varphi_{X_1}(\lambda)-1)\}}$
Binomiale Négative $(\alpha, p)$	$\alpha \in ]0, 1[$ $p > 0$	$\frac{\Gamma(p+n)}{\Gamma(n+1)\Gamma(p)} \alpha^n (1-\alpha)^p$	$\left(\frac{1-\alpha}{1-\alpha\varphi_{X_1}(\lambda)}\right)^p$
Logarithmique $(\theta)$	$\theta \in ]0, 1[$	$\theta^n / (n \ln(1-\theta)^{-1})$	$\frac{\ln(1-\theta\varphi_{X_1}(\lambda))}{\ln(1-\theta)}$

TAB. 3.2

Considérons maintenant la variable aléatoire  $T$  de densité  $m^{-1}\mathbb{P}(S_N > x)$  où  $m = \mathbb{E}[S_N] = \mathbb{E}[N] \mathbb{E}[X_1]$ . Sa fonction génératrice des moments est donnée par :

$$\varphi_T(\lambda) = \frac{\varphi_{S_N}(\lambda) - 1}{\lambda m}.$$

On peut alors générer une famille exponentielle de variables aléatoires  $Y_\lambda$  de fonction de distribution  $G_\lambda$  et densité  $g_\lambda$  donnée par :

$$g_\lambda(x) = \frac{\lambda e^{\lambda x} \mathbb{P}[S_N > x]}{\varphi_{S_N}(\lambda) - 1}, \quad \lambda \in \Lambda_T.$$

Posons maintenant  $q(\lambda) = \ln \varphi_{S_N}(\lambda)$ ,  $m(\lambda) = q'(\lambda)$  et  $s^2(\lambda) = q''(\lambda)$ , nous avons

$$\begin{aligned} a(\lambda) &= \mathbb{E}[Y_\lambda] = \frac{\varphi_{S_N}(\lambda)}{\varphi_{S_N}(\lambda) - 1} m(\lambda) - \frac{1}{\lambda}, \\ b^2(\lambda) &= \mathbb{V}[Y_\lambda] = \left( \frac{\varphi_{S_N}(\lambda)}{\varphi_{S_N}(\lambda) - 1} \right)^2 s^2(\lambda) + \frac{1}{\lambda^2} - \frac{\varphi_{S_N}''(\lambda)}{(\varphi_{S_N}(\lambda) - 1)^2}. \end{aligned}$$

Soit  $\lambda_{T,\infty} = \sup \Lambda_T$ . Balkema, Klüppelberg et Resnick ([BKR99] et [BKR03]) ont montré que les éléments de la famille peuvent être normalisés de telle manière à ce que

$$U_\lambda := \frac{Y_\lambda - a(\lambda)}{b(\lambda)} \xrightarrow{d} Z, \quad \text{lorsque } \lambda \rightarrow \lambda_{T,\infty}.$$

Une condition nécessaire est que  $\lambda_{T,\infty}$  ne soit pas dans  $\Lambda_T$ .

**Proposition 3.3.** [BKR03]

i) Si  $U_\lambda \xrightarrow{d} Z$ , alors il existe  $\beta \in \mathbb{R}$  tel que  $\varphi_Z = \varphi_\beta$  où

$$\ln \varphi_\beta(\lambda) = \begin{cases} -\beta^{-2} \ln(1 - \lambda\beta) - \lambda/\beta & \text{si } \beta \neq 0, \quad \lambda\beta < 1. \\ \lambda^2/2 & \text{si } \beta = 0. \end{cases}$$

i.e.  $Z$  a une distribution Gaussienne standard ou  $Z$  (ou  $-Z$ ) a une distribution Gamma normalisée.

ii) Une condition nécessaire et suffisante lorsque  $\beta \neq 0$  est que  $\lambda_{T,\infty}$  est fini et

$$\lim_{\lambda \rightarrow \lambda_{T,\infty}} (\lambda_{T,\infty} - \lambda) \mathbb{E}[Y_\lambda] = 1/\beta^2.$$

Nous pouvons alors en déduire la proposition suivante.

**Proposition 3.4.** Si  $(Y_\lambda - b(\lambda)) / a(\lambda) \xrightarrow{d} Z_\beta$ , alors

$$\begin{aligned} \mathbb{P}(S_N > x) &\underset{x \rightarrow \infty}{\sim} \frac{\beta^{-2(\beta^{-2}-1/2)} e^{-\beta^{-2}} \exp(-xb^{-1}(x)) \varphi_{S_N}(b^{-1}(x))}{\Gamma(\beta^{-2}) b^{-1}(x) a(b^{-1}(x))}, & \text{si } \beta \neq 0, \\ \mathbb{P}(S_N > x) &\underset{x \rightarrow \infty}{\sim} \frac{\exp(-xb^{-1}(x)) \varphi_{S_N}(b^{-1}(x))}{(2\pi)^{1/2} b^{-1}(x) a(b^{-1}(x))}, & \text{si } \beta = 0. \end{aligned}$$

Si l'on considère les exemples donnés dans le tableau précédent, les distributions asymptotiques sont alors données par

Distribution de $N$	$\beta$	$\ln \varphi_\beta(t)$
Binomiale $(p, m)$	0	$t^2/2$
Poisson $(\mu)$	0	$t^2/2$
Binomiale Négative $(\alpha, p)$	$1/p^{1/2}$	$-p \ln(1 - t/p^{1/2}) - p^{1/2}t$
Logarithmique $(\theta)$	non défini	ne converge pas

TAB. 3.3

Considérons l'exemple suivant :  $N$  a une distribution Binomiale Négative  $(\alpha, p)$  et il existe une constante  $\lambda_\infty > 0$  telle que  $\varphi_{X_1}(\lambda_\infty) = \alpha^{-1}$ . Nous avons alors pour  $\lambda < \lambda_\infty$

$$\varphi_{S_N}(\lambda) = \left( \frac{1 - \alpha}{1 - \alpha \varphi_{X_1}(\lambda)} \right)^p$$

et  $\lambda_{T,\infty} = \lambda_\infty$ . Par conséquent,

$$m(\lambda) = \frac{\alpha p \varphi_{X_1}'(\lambda)}{1 - \alpha \varphi_{X_1}(\lambda)} \quad \text{et} \quad b^{-1}(x) = \lambda_\infty - \frac{p}{x} + o\left(\frac{1}{x}\right).$$

Avec  $\beta = 1/p^{1/2}$ , nous retrouvons les résultats de [EMT85]:

$$\mathbb{P}(S_N > x) \underset{x \rightarrow \infty}{\sim} \frac{1}{\lambda_\infty \Gamma(p)} \left( \frac{(1-\alpha)}{\alpha \varphi'_{X_1}(\lambda_\infty)} \right)^p x^{p-1} \exp(-\lambda_\infty x).$$

### **$N$ a une queue de distribution fine et $X_1$ a une queue de distribution épaisse**

Il existe un résultat bien connu lorsque  $X_1$  a une distribution sous-exponentielle (voir par exemple [EV82], [EV88] et le théorème A3.20 dans [EKM97]).

**Proposition 3.5.** *Si  $F_X \in \mathcal{S}$ , alors*

$$\mathbb{P}(S_N > x) \underset{x \rightarrow \infty}{\sim} \mathbb{E}[N] \mathbb{P}(X_1 > x).$$

### **3.4.2 $N$ a une queue de distribution épaisse**

Avec Johan Segers, nous avons décidé de nous intéresser dans [A11] au cas où  $N$  a une queue de distribution épaisse. Il n'existe pas de réponse générale (voir par exemple [Sc99] et [St73]). Nous avons étudié quelques cas particuliers pour lesquels la queue de distribution de  $N$  est plus épaisse que celle de  $X_1$ .

#### **$N$ a une queue de distribution à variations consistantes**

Nous avons établi la proposition suivante.

**Proposition 3.6.** *Si la queue de distribution de  $N$  est à variations consistantes, si  $\mathbb{E}[X_1^r] < \infty$  pour un  $r > 1$  et si l'une des deux conditions suivantes est satisfaite:*

(i) *cas  $\mathbb{E}[N] < \infty$ :*

$$\mathbb{P}(X_1 > x) = o(\mathbb{P}(N > x)), \quad x \rightarrow \infty;$$

(ii) *cas  $\mathbb{E}[N] = \infty$ : il existe  $q$  avec  $1 \leq q < r$  tel que*

$$\limsup_{x \rightarrow \infty} \frac{\mathbb{E}[N \mathbf{1}\{N \leq x\}]}{x^q \mathbb{P}(N > x)} < \infty;$$

alors

$$\mathbb{P}(S_N > x) \underset{x \rightarrow \infty}{\sim} \mathbb{P}(\mathbb{E}[X_1] N > x).$$

#### **$N$ est dans le domaine d'attraction de la loi de Gumbel**

Rappelons tout d'abord que  $N$  est dans le domaine d'attraction de la loi de Gumbel si et seulement si il existe une fonction  $a$  définie sur un voisinage de l'infini telle que

$$\lim_{x \rightarrow \infty} \frac{\mathbb{P}(N > x + ya(x))}{\mathbb{P}(N > x)} = e^{-y}.$$

**Proposition 3.7.** *Si  $X_1$  est à queue fine et si  $N$  est dans le domaine d'attraction de la loi de Gumbel tel que*

$$\liminf_{x \rightarrow \infty} \frac{\ln a(x)}{\ln x} > \frac{1}{2},$$

alors

$$\mathbb{P}(S_N > x) \underset{x \rightarrow \infty}{\sim} \mathbb{P}(\mathbb{E}[X_1] N > x).$$

Dans le cas où

$$\liminf_{x \rightarrow \infty} \frac{\ln a(x)}{\ln x} \leq \frac{1}{2},$$

le résultat précédent n'est plus vrai. Supposons par exemple que  $X_1$  a une distribution exponentielle de paramètre 1. Pour  $x \geq 0$ , la distribution de la variable  $Z_x = \max\{n = 0, 1, \dots : S_n \leq x\}$  est une distribution de Poissons de moyenne  $x$ , et

$$\mathbb{P}(S_N > x) = \mathbb{P}(N > Z_x) = \mathbb{E}[\exp\{-g(Z_x)\}],$$

où  $g(z) := -\ln \mathbb{P}(N > z)$ . Le comportement asymptotique de  $\mathbb{E}[\exp\{-g(Z_x)\}]$ , lorsque  $x$  tend vers l'infini, a été étudié dans [FK00] dans le cas général où  $Z_x$  est une somme de variables aléatoires positives indépendantes et équidistribuées. Si l'on considère le cas  $N = \lfloor Y \rfloor$  avec  $\mathbb{P}(Y > y) = \exp(-y^\beta)$  pour  $y \geq 0$  où  $\beta \in [1/2, 2/3[$ , alors on obtient

$$\mathbb{P}(S_N > x) \underset{x \rightarrow \infty}{\sim} \mathbb{P}(N > x) \exp(\beta^2 x^{2\beta-1}/2),$$

et on s'aperçoit que  $\mathbb{P}(S_N > x)$  n'est plus asymptotiquement équivalente à  $\mathbb{P}(N > x)$ .

### 3.5 Conclusion et projets de recherche

Les distributions à queue épaisse sont des objets qui intéressent beaucoup les assureurs car elles sont omniprésentes dans leur quotidien. AXA a notamment créé récemment une chaire "Assurance et Risques Majeurs" qui étudie en particulier l'assurabilité des grands risques. De nombreux cas de la dernière section restent encore à étudier. Nous projetons d'en regarder quelques un dans les années à venir.



## Chapitre 4

# Statistique des processus et séries financières haute fréquence

### 4.1 Introduction

#### 4.1.1 Séries financières haute fréquence et bruit de microstructure

Les séries de données enregistrées en temps réel sur les marchés financiers sont dénommées “séries financières haute fréquence” par opposition aux séries échantillonnées quotidiennement. Les prix de transaction, les prix de cotation et les cinq meilleurs limites (à l’achat ou à la vente) sont des données désormais disponibles à l’échelle de la seconde pour un grand nombre d’actifs. Les praticiens ont en effet développé des outils informatiques capables d’enregistrer “à la volée” des milliers d’opérations.

D’une part, il est bien connu que ces données haute fréquence présentent des caractéristiques atypiques telles que des modifications non équi-réparties dans le temps, des échelles de valeurs discrètes, des saisonnalités très fortes à l’intérieur des journées de marché avec notamment des comportements très différents à l’ouverture ou à la clôture (voir par exemple [En00] et [GJ01]). D’autre part, une grande partie des mathématiques financières est basée sur l’hypothèse que les dynamiques des prix sont des semi-martingales d’Itô. Ainsi un problème important est de construire un modèle capable de réconcilier théorie et pratique. En effet, autant il est raisonnable d’admettre que les données quotidiennes, mensuelles, ... peuvent être considérées comme des réalisations de semi-martingales continues, autant il est incorrect de le faire pour des séries à l’échelle de la seconde où la discrétisation imposée par la grille de prix du marché domine. Il est plutôt usuel de considérer que les prix de transaction ou de cotation sont des réalisations bruitées du vrai prix (celui de la semi-martingale) que nous qualifierons par la suite de prix “efficient” et d’utiliser une modélisation *ad hoc* pour ce bruit. Nous préférons ici construire un modèle qui permette de passer de la semi-martingale aux prix observés et ainsi d’obtenir un bruit aux propriétés plus satisfaisantes que celles imposées dans l’approche traditionnelle.

Considérons un actif de prix  $X_t$  à la date  $t \in [0,1]$  (une journée de cotation) et dont la dynamique est régie par une semi-martingale continue d’Itô. Posons  $t_0 = 0$  et notons  $P_0$  le prix d’ouverture. Pour  $i \geq 1$ , soit  $t_i$  le  $i$ -ième temps où se produit un changement du prix de transaction de l’actif et soit  $P_{t_i}$  le prix associé. Nous définissons le processus “du dernier prix traité”  $(P_t)_{t \geq 0}$  comme le processus càdlàg constant par morceaux construit à partir des couples  $(t_i, P_{t_i})_{i \geq 0}$ . La différence

$$\varepsilon_t = \log P_t - \log X_t \tag{4.1}$$

définit le “bruit de microstructure” à la date  $t$ . Ce bruit regroupe plusieurs types d’imperfections dues aux caractéristiques du marché: processus d’appariement, prix discrets, effets informationnels

(asymétries d'informations, diffusion d'informations privilégiées,...).

Avant de discuter plus précisément des diverses modélisations du bruit de microstructure qui ont été introduites dans la littérature, considérons la variation quadratique de  $(\log X_t)_{t \geq 0}$  et la covariation entre  $(\varepsilon_t)_{t \geq 0}$  et  $(\log X_t)_{t \geq 0}$ . Supposons qu'elles soient bien définies, elles sont alors données par les formules respectives

$$\begin{aligned} [\log X]_t &= \mathbb{P}\text{-}\lim_{n \rightarrow \infty} \sum_{j=1}^n (\log X_{s_{j,n}} - \log X_{s_{j-1,n}})^2, \\ [\varepsilon, \log X]_t &= \mathbb{P}\text{-}\lim_{n \rightarrow \infty} \sum_{j=1}^n (\varepsilon_{s_{j,n}} - \varepsilon_{s_{j-1,n}})(\log X_{s_{j,n}} - \log X_{s_{j-1,n}}), \end{aligned}$$

pour n'importe quelle suite de partition déterministe ( $0 = s_{0,n} < s_{1,n} < \dots < s_{n,n} = t$ ) telle que  $\sup_j \{s_{j+1,n} - s_{j,n}\}$  tend vers zéro quand  $n$  tend vers l'infini. Ces deux quantités sont des objets très utiles pour déduire certaines propriétés élémentaires du bruit de microstructure. Si l'on suppose que  $(P_t)_{t \geq 0}$  un processus de sauts pour lequel le nombre de sauts est presque sûrement fini sur tout intervalle borné, nous en déduisons que

$$[\log P, \log X]_t = [\log X]_t + [\varepsilon, \log X]_t = 0$$

et

$$[\varepsilon]_t = [\log P]_t - [\log X]_t - 2[\varepsilon, \log X]_t = [\log P]_t + [\log X]_t.$$

Par conséquent

- (i)  $[\varepsilon]_t$  is presque sûrement fini,
- (ii)  $[\varepsilon, \log X]_t = -[\log X]_t < 0$  (ce qui signifie grossièrement que les incréments du bruit de microstructure sont négativement corrélés avec les rendements du prix efficient),
- (iii) si  $[\log X]_t$  et  $[\log P]_t$  présentent des saisonnalités, c'est aussi le cas pour  $[\varepsilon]_t$ .

Un premier panorama des différents modèles de bruit de microstructure rencontrés dans la littérature financière est donné dans [BSh07]. Nous nous restreignons ici aux différentes familles de modèles qui sont cohérents avec l'hypothèse d'une semi-martingale continue.

La première famille considère un bruit de microstructure qui est stationnaire dans le temps d'observation. Les exemples les plus utilisés sont le bruit blanc (voir par exemple [Zh96] ou [ZMA05]) et le bruit avec covariances stationnaires ne dépendant pas de la fréquence d'échantillonnage considérée (voir par exemple [BHLS09]). De plus, le bruit de microstructure est supposé indépendant du prix efficient. Cet ensemble d'hypothèses est pratique pour mener des calculs et n'est pas trop irréaliste lorsque l'on considère des fréquences d'échantillonnage correspondant à des périodes plus petites que cinq minutes. Cependant, il est mathématiquement non satisfaisant car il est en contradiction avec les trois propriétés du bruit de microstructure mentionnées plus haut. En effet, aux très grandes fréquences, les erreurs dues à la discrétisation des prix et aux effets de la saisonnalité conduisent à des dépendances non-linéaires très fortes entre le bruit de microstructure et le prix efficient et à une hétéroscédasticité très complexe du bruit. Ainsi, les hypothèses précédentes sont beaucoup trop fortes lorsque l'on considère des données à très haute fréquence. Notons qu'un modèle avec un bruit hétéroscédastique qui est corrélé avec le prix efficient et dont la variance dépend de la fréquence d'échantillonnage a été récemment proposé dans [KL08] (voir aussi [BHLS09]).

La seconde famille considère que le bruit de microstructure est une diffusion dans le temps calendaire et qu'il est indépendant du prix efficient. Ce cas a été étudié par exemple dans [HL06]. Contrairement au cas précédent, il satisfait la condition  $[\varepsilon]_t < \infty$  p.s.. Cependant, il est clair qu'il ne peut pas prendre en compte l'aspect discret des prix.

La troisième famille a été introduite dans [LM07]. Le prix efficient est “contaminé” par l’intermédiaire d’un noyau Markovien. Plus précisément, les auteurs supposent qu’il existe une distribution conditionnelle  $K(x,dp)$  telle que, conditionnellement à  $X_{t_i}$ , la loi de  $P_{t_i}$  est donnée par

$$\mathbb{P}_{X_{t_i}}(P_{t_i} \leq p) = K(X_{t_i}, p).$$

De plus, ils supposent que, conditionnellement à  $(X_t)_{t \geq 0}$ ,  $P_{t_0}, P_{t_1}, \dots$  sont indépendants. Cette approche est aussi développée dans [JLMPV09]. Dans la mesure où le noyau ne dépend pas de la fréquence d’échantillonnage, ce type de friction implique que, dans la plupart des cas,  $[\varepsilon]_t = \infty$ .

La quatrième famille considère des modèles structurels de bruit de microstructure se basant sur des procédures de discrétisation. Le cas de l’approximation par partie entière est abordé dans [DJ97] et [Ro07a]. Ce cas suppose que  $P_t$  est égal à  $X_t^{(\alpha)}$  à certaines dates,  $\alpha$  étant la taille du tick de cotation (c’est-à-dire le plus petit écart de prix possible) et  $s^{(\alpha)} = \alpha \lfloor s/\alpha \rfloor$  étant la valeur appartenant à la grille de prix et la plus proche de  $s$ . Cependant, s’il l’on suppose de plus que cette égalité est vraie pour toute date  $t$ , alors  $(P_t)_{t \geq 0}$  devient un processus de saut avec une activité infinie, ce qui en pratique n’est pas satisfaisant. Dans [Fu09], il est supposé que le prix de transaction saute vers le haut ou vers le bas d’un tick dès que le prix efficient croise la grille de prix à un prix qui diffère de l’ancien prix de transaction. Les limites de ce modèle sont d’abord qu’il n’est pas capable de reproduire la fameuse danse de la fourchette (autocorrélation d’ordre 1 négative des prix de transaction) et que, d’autre part, les délais entre la date de franchissement et la date où la transaction a lieu (délai de réaction et de passage d’ordre) est nul. Pour remédier à la première limite, on peut considérer un modèle où le prix de transaction saute d’un tick dès lors que la distance entre le prix efficient le dernier prix traité est égale à une constante qui est plus grande que la moitié du tick. Un tel modèle est abordé dans [La07]. Cependant, il ne peut être utilisé que pour des actifs très liquides pour lesquels les variations de prix sont toujours d’un tick, ce qui est irréaliste pour une très grande majorité des actifs. Ce type de modèle a aussi été proposé sous une version en temps discret dans [Ro07a]. Sa finalité était de reproduire les autocorrélations du bruit dans les très hautes fréquences.

L’article [A13] propose un modèle de bruit de microstructure de la quatrième famille qui reproduit l’ensemble des faits stylisés caractéristiques des données de transaction à très haute fréquence (prix discrets, durées entre transactions aléatoires, fortes saisonnalités) avec l’hypothèse d’un prix efficient régi par une semi-martingale d’Itô continue. En particulier, le modèle proposé permet d’avoir des prix discrets, des variations de prix de un ou plusieurs ticks et un comportement dépendant de plusieurs facteurs tels que, par exemple, le carnet d’ordre. Dans notre modèle, le bruit de microstructure est structurel, c’est-à-dire qu’il est issu du processus de prix efficient et des anticipations des acteurs du marché. Enfin, contrairement aux précédents articles, nous ne supposons pas que les délais de réaction des acteurs du marché sont nuls.

#### 4.1.2 Estimation de la volatilité intégrée en présence de bruit de microstructure

La contribution du papier [A12] est de proposer des estimateurs de la volatilité intégrée (c’est-à-dire la variation quadratique de  $(\log X_t)_{t \geq 0}$ ) et de la covolatilité intégrée. Concentrons nous dans un premier temps sur l’estimation de la volatilité intégrée. Dans un contexte sans bruit de microstructure, l’estimateur naturel, disposant depuis de nombreuses années d’une théorie asymptotique aboutie, est donné par la volatilité réalisée qui est la somme des variations au carré des logarithmes du prix efficient obtenues sur une subdivision de l’intervalle de temps considéré (voir par exemple [BSh02], [Ja06]). En pratique, cet estimateur est raisonnable pour des fréquences d’échantillonnage correspondant à des périodes inférieures à 15 minutes. Aux plus hautes fréquences, la présence du bruit de microstructure rend l’estimateur non convergent. Ainsi de nombreuses méthodes dépendant du type de bruit de microstructure ont été développées pour éliminer l’effet du bruit dans l’estimation de la volatilité intégrée. Elles peuvent reposer sur des procédures d’échantillonnage à plusieurs

échelles afin d'estimer la variabilité du bruit de microstructure (voir par exemple [ZMA05], [KL08], [Z06a]), sur des méthodes basées sur des noyaux qui permettent d'éliminer le bruit de microstructure par la prise en compte des autocorrélations du processus (voir par exemple [Zh96], [BHLS09]), sur des méthodes par ondelettes (voir [Ro07b]), ou sur des procédures d'échantillonnage endogènes basées sur le temps de marché ou le temps des variations de prix (voir par exemple [CF88], [Ba88], [Fu09]).

L'approche du papier est basée sur un échantillonnage dans le temps des variations de prix. Notre méthode consiste à estimer dans un premier temps les valeurs du prix efficient à des temps d'arrêt, puis à calculer la volatilité réalisée à partir de ces valeurs. Nous montrons que notre estimateur est convergent et nous donnons la théorie asymptotique lorsque la taille du tick tend vers zéro. Notons que, contrairement à d'autres auteurs, nous étudions la variation quadratique de  $(\log X_t)_{t \geq 0}$  et non pas celle de  $(X_t)_{t \geq 0}$ . Bien que ces deux quantités soient proches (à un facteur d'échelle près), la première quantité est plus pertinente d'un point de vue financier car elle apparaît naturellement dans les procédures de tarification des produits dérivés. Par contre, l'estimation de la variation quadratique de  $(\log X_t)_{t \geq 0}$  est plus compliquée dans le cas de prix discrets.

Nous fournissons aussi un estimateur convergent de la covolatilité intégrée dans le cas où deux actifs sont observés. L'estimation de cette quantité a été moins étudiée que la volatilité intégrée (voir [Z06b], [VL07]). En effet, en plus du bruit de microstructure, le problème de l'asynchronicité des données (les variations de prix des deux actifs ne sont pas observées aux mêmes instants) vient se greffer. Notre méthode consiste à utiliser les valeurs du prix efficient et puis à adapter la technique d'estimation introduite dans [HY05].

### 4.1.3 Erreurs de couverture en présence de bruit de microstructure

Dans [A14], nous nous intéressons à l'erreur de couverture d'un produit dérivé européen dans le cadre de notre modèle avec zones d'incertitude. En mathématiques financières, la tarification et la couverture des produits dérivés ont été initialement étudiées dans un contexte de marché "sans friction". Les hypothèses d'un tel marché sont les suivantes :

- Il est possible d'emprunter et de placer de l'argent à un taux unique sans risque.
- Le prix de transaction est égal au prix efficient, quelque soit le volume demandé et son signe (à l'achat ou à la vente).
- Il est possible d'acheter ou de vendre de manière instantanée et continue.
- Il n'y a pas de coûts de transaction.
- L'actif est parfaitement divisible (il est possible d'acheter ou vendre n'importe quelle quantité).
- Les ventes à découvert sont autorisées.

Le fait que l'une de ces hypothèses ne soit pas vérifiée rend beaucoup plus complexe la détermination de la stratégie de couverture. Le cas de la restriction des ventes à découvert a été étudié par exemple dans [JK95] et la présence de coûts de liquidité dans [CST09]. La conséquence des coûts de transaction couplés avec des stratégies de couverture comportant un nombre fini de réallocations des actifs du portefeuille a aussi été étudiée de manière approfondie (voir par exemple [KS97], [KPT99]).

Dans [A14], nous considérons notre modèle avec zones d'incertitude qui conduit à une modélisation des variations de prix et des durées entre transaction relativement réaliste et contraignons les prix de transaction à appartenir à une grille de prix. L'aspect discret des prix a deux conséquences importantes. La première est l'impossibilité d'acheter ou de vendre l'actif risqué considéré au prix efficient : le bruit de microstructure conduit à un coût qui ne peut être éliminé complètement. La seconde vient du fait que le prix de transaction change un nombre fini de fois sur un intervalle de temps borné et qu'il est raisonnable de supposer qu'une nouvelle recombinaison du portefeuille ne

sera mis en application que lorsque le prix de transaction aura changé. Ainsi nous remettons en cause la deuxième hypothèse d'un marché sans friction et donc par conséquent la troisième devient inappropriée.

Pour analyser l'erreur de couverture due au bruit de microstructure, nous considérons un agent qui adapte une stratégie usuelle utilisée pour un marché sans friction. Cette stratégie usuelle est donnée par un portefeuille de réplication basé sur une hypothèse (qui peut être fausse) d'un modèle à volatilité locale (voir en particulier [EJS98]). Nous considérons deux stratégies de couverture dérivées de la stratégie en temps continu en l'adaptant à des réallocations à des instants aléatoires : (i) la composition du portefeuille de couverture est remise à jour chaque fois que le prix de transaction change, (ii), la composition du portefeuille de couverture est remise à jour uniquement lorsque le prix de transaction a varié d'au moins une certaine valeur. Nous développons une approche asymptotique dans l'esprit de [BKL00], [HM05] et [TV08].

Nous montrons que, lorsque le nombre de transactions tend vers l'infini, l'erreur de couverture ne tend pas vers zéro dans le cas de la première stratégie à cause de la présence du bruit de microstructure. Par contre, cette erreur peut être considérablement réduite dans le cas d'une stratégie du second type.

## 4.2 Le modèle avec zones d'incertitude

### 4.2.1 Description du modèle

Dans cette section, nous expliquons la construction de notre modèle. Dans un cadre idéal où le prix efficient serait observé, les participants du marché échangeraient lorsque le prix efficient croiserait la grille de prix. En pratique, il y a toujours une incertitude sur ce prix efficient. Ainsi, il y a une modification du prix de transaction seulement si les acheteurs et les vendeurs sont convaincus que le prix efficient est suffisamment loin du dernier prix traité. Nous introduisons un paramètre  $\eta$  destiné à quantifier l'aversion aux changements de prix des participants du marché et proposons un modèle qui prend en compte cette aversion ( $\eta$  est une proportion du tick de cotation).

Nous notons  $Y_t$  le logarithme du prix efficient de l'actif ( $Y_t = \log X_t$ ). Sur un espace de probabilité suffisamment riche et muni d'une filtration  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ , nous supposons que  $(Y_t)_{t \geq 0}$  est une semi-martingale  $\mathcal{F}_t$ -adaptée, continue de la forme

$$Y_t = \log X_0 + \int_0^t a_u du + \int_0^t \sigma_u - dW_u,$$

où  $(W_t)_{t \geq 0}$  est un  $\mathcal{F}$ -mouvement Brownien standard,  $(a_t)_{t \geq 0}$  est un processus prévisible à trajectoires localement bornées et  $(\sigma_t)_{t \geq 0}$  est un processus positif  $\mathcal{F}_t$ -adapté avec des trajectoires càdlàg.

La grille des prix de transaction est définie par  $\{k\alpha; k \in \mathbb{N}\}$ , où  $\alpha$  est la taille du tick. Pour  $k \in \mathbb{N}$  et  $0 < \eta < 1$ , nous définissons la zone  $U_k$  par  $U_k = [0, \infty) \times (d_k, u_k)$  avec

$$d_k = (k + 1/2 - \eta)\alpha \text{ et } u_k = (k + 1/2 + \eta)\alpha.$$

Ainsi,  $U_k$  est une bande centrée sur les valeurs intermédiaires  $(k + 1/2)\alpha$  de la grille, voir la figure 4.1. Notons que si  $\eta$  est plus petit que  $1/2$ , il n'y a pas de recouvrement entre les bandes  $U_k$ .

Nous supposons que le prix de transaction peut sauter du prix  $k'\alpha$  au prix  $k\alpha$  avec  $k' \neq k$  seulement une fois que le prix efficient est sorti de la zone  $U_k$  par le bas ou est sorti de la zone  $U_{k-1}$  par le haut et pourvu que les conditions sur le marché soient favorables pour qu'un échange ait lieu (au moins un acheteur et un vendeur). D'une certaine façon, le prix de transaction change seulement si le prix efficient est près d'un nouveau multiple de  $\alpha$  et les participants du marché veulent échanger. Les zones  $(U_k)_{k \in \mathbb{N}}$  représentent des bandes à l'intérieur desquelles le prix efficient

ne peut conduire à une modification du prix de transaction. Par conséquent, nous les dénommerons zones d'incertitude.

Précisons maintenant la construction de la suite des temps de sortie des zones d'incertitude qui produiront un changement du prix de transaction  $(\tau_i)_{i \geq 0}$ . Posons  $\tau_0 = 0$  et supposons, sans perte de généralité, que  $\tau_1$  est le temps de sortie de  $(X_t)_{t \geq 0}$  de l'ensemble  $(d_{k_0-1}, u_{k_0})$  où  $k_0 = X_0^{(\alpha)}$ . Nous introduisons ensuite une suite de variables aléatoires discrètes  $(L_i)_{i \geq 1}$   $\mathcal{F}_{\tau_i}$ -mesurables qui seront les valeurs absolues (en nombre de ticks) des sauts du prix de transaction entre la  $i$ -ième et la  $(i+1)$ -ième transaction conduisant à un changement de prix. Comme nous l'expliquerons plus tard, les distributions de ces variables dépendront de quantités du marché définies à l'instant  $\tau_i$  comme par exemple des quantités reliées aux volumes traités, à la fourchette de prix entre les meilleurs prix d'achat ou de vente, ... Nous pouvons alors définir de manière récursive  $\tau_{i+1}$  comme le temps de sortie de  $(X_t)_{t > \tau_i}$  de l'ensemble  $(d_{k_i-L_i}, u_{k_i+L_i-1})$ , où  $k_i = X_{\tau_i}^{(\alpha)}$ , c'est-à-dire,

$$\tau_{i+1} = \inf \left\{ t : t > \tau_i, X_t = X_{\tau_i}^{(\alpha)} - \alpha(L_i - \frac{1}{2} + \eta) \text{ ou } X_t = X_{\tau_i}^{(\alpha)} + \alpha(L_i - \frac{1}{2} + \eta) \right\}. \quad (4.2)$$

En particulier, si  $X_{\tau_i} = d_j$  pour un certain  $j \in \mathbb{N}$ ,  $\tau_{i+1}$  est le temps de sortie de  $(X_t)_{t > \tau_i}$  de l'ensemble  $(d_{j-L_i}, u_{j+L_i-1})$ , et si  $X_{\tau_i} = u_j$  pour un certain  $j \in \mathbb{N}$ ,  $\tau_{i+1}$  est le temps de sortie de  $(X_t)_{t > \tau_i}$  de l'ensemble  $(d_{j-L_i+1}, u_{j+L_i})$  (voir la figure 4.1).

Rappelons maintenant que le processus du dernier prix traité est caractérisé par les couples des dates de transaction et des prix de transaction  $(t_i, P_{t_i})_{i \geq 0}$ . Soit  $P_0 = X_0^{(\alpha)}$ . Nous supposons que, pour tout  $i \geq 1$ ,

$$\tau_i \leq t_i < \tau_{i+1} \text{ et } P_{t_i} = X_{\tau_i}^{(\alpha)}.$$

Cela signifie que, entre  $\tau_i$  et  $\tau_{i+1}$ , au moins une transaction a eu lieu au prix  $P_{t_i}$  et  $t_i$  est le temps de la première transaction à ce prix. Il est possible de modéliser l'occurrence d'arrivée de cette première transaction à l'aide d'un processus à intensité où l'intensité deviendrait infinie en s'approchant des nouvelles barrières associées à  $\tau_{i+1}$ . La différence  $t_i - \tau_i$  peut être considérée comme le délai dû au temps de réaction des participants du marché et/ou du processus d'échange. Notons que  $\alpha L_i$  est la valeur absolue du saut du prix de transaction entre la  $i$ -ième et la  $(i+1)$ -ième transaction et que

$$P_{t_i} = X_{\tau_i} + \text{sign}(X_{\tau_i} - X_{\tau_{i-1}})(1/2 - \eta)\alpha = X_{\tau_i} + \text{sign}(P_{t_i} - P_{t_{i-1}})(1/2 - \eta)\alpha. \quad (4.3)$$

Ainsi, si l'on connaît (ou estime)  $\eta$ , on peut retrouver (ou estimer)  $X_{\tau_i}$  à partir de  $P_{t_{i-1}}$  et  $P_{t_i}$ . La figure 4.1 donne un exemple de trajectoire dans un cas où  $0 < \eta < 1/2$ .

Finalement, nous caractérisons la distribution conditionnelle des sauts des prix de transaction en ticks. Nous supposons que la taille des sauts est bornée (ce qui empiriquement n'est pas restrictif) et notons  $m$  leur valeur maximale. Pour  $k = 1, \dots, m$  et  $t > 0$ , posons

$$N_{\alpha, t, k}^{(a)} = \sum_{\tau_i \leq t} 1\{|X_{\tau_i} - X_{\tau_{i-1}}| = \alpha(k-1+2\eta)\} \text{ et } N_{\alpha, t, k}^{(c)} = \sum_{\tau_i \leq t} 1\{|X_{\tau_i} - X_{\tau_{i-1}}| = \alpha k\}$$

le nombre "d'alternances" de direction de prix et respectivement le nombre de "continuations" de direction de prix. Une alternance (resp. une continuation) de  $k$  ticks est un saut de  $k$  ticks dont la direction est opposée (resp. dans le même sens) que le précédent saut (voir la figure 4.1). Remarquons que pour des petites (resp. grandes) valeurs de  $\eta$ , on observera un grand nombre d'alternances (resp. de continuations). Soit  $(\chi_t)_{t \geq 0}$  un processus continu de dimension  $m$ ,  $\mathcal{F}_t$ -adapté. Nous définissons la filtration  $\mathcal{E}$  par

$$\mathcal{E}_t = \cap_{s > t} \sigma(X_r, \chi_r, N_{\alpha, r, k}^{(a)}, N_{\alpha, r, k}^{(c)}, k = 1, \dots, m, r \leq s).$$

Nous supposons que, conditionnellement à  $\mathcal{E}_{\tau_i}$ ,  $L_i$  est une variable aléatoire discrète sur  $\llbracket 1, m \rrbracket$  satisfaisant

$$\mathbb{P}_{\mathcal{E}_{\tau_i}}(L_i = k) = p_k(\chi_{\tau_i}), \quad 1 \leq k \leq m, \quad (4.4)$$

pour des fonctions  $p_k$  positives, avec une dérivée bornée. En pratique,  $\chi_t$  peut représenter des quantités reliées au volume traité, à la fourchette de prix entre les meilleurs prix d'achat ou de vente, ou encore les volumes proposés aux meilleurs prix d'achat ou de vente.

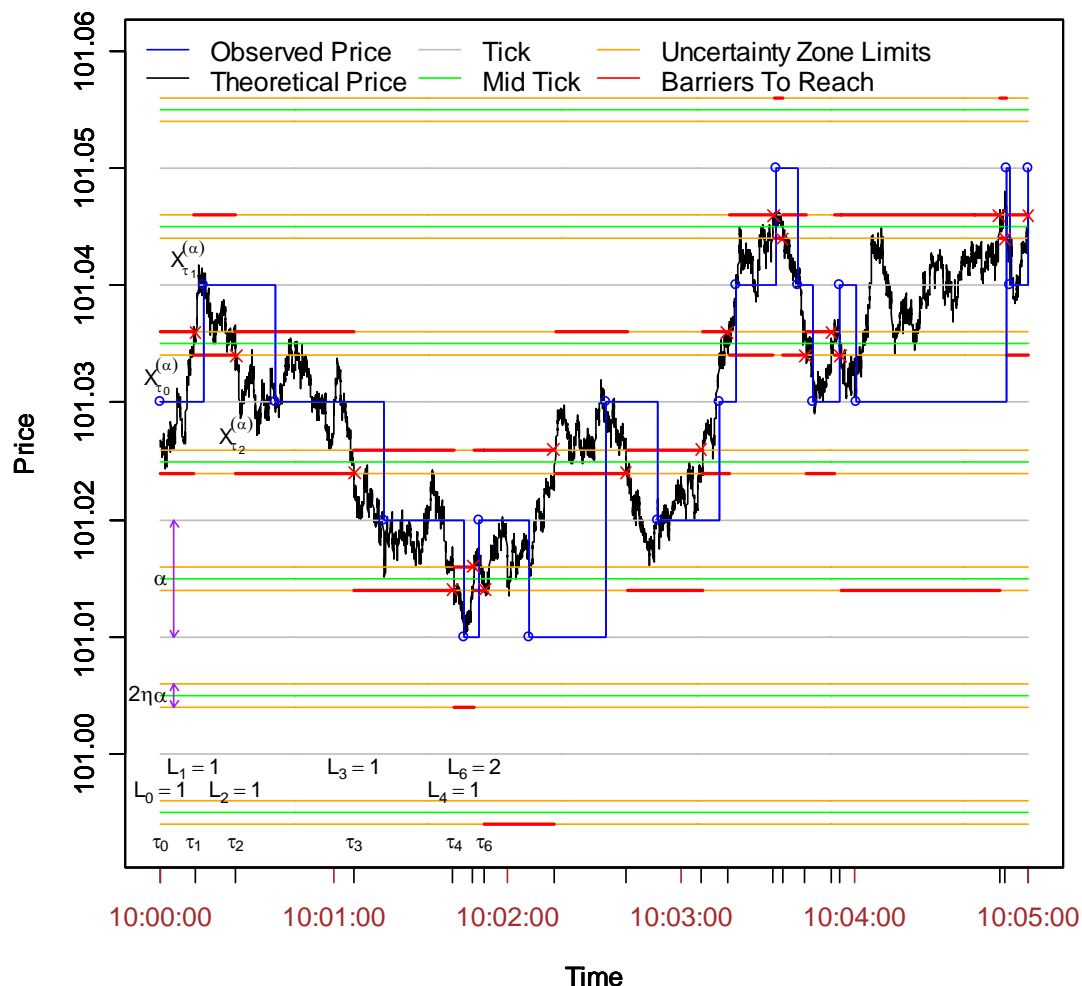


FIG. 4.1 – Un exemple de trajectoire du prix efficient et du prix de transaction.

#### 4.2.2 Discussion

Notre modèle est compatible avec l'hypothèse d'une semi-martingale continue d'Itô. De plus, notre processus de bruit de microstructure vérifie les propriétés essentielles présentées dans la section 4.1.1. Les prix de transaction prennent des valeurs sur la grille de prix imposée par l'organisation du marché et le prix de transaction change à des instants aléatoires. Les prix peuvent changer de un ou plusieurs ticks, ce qui est le cas pour la majorité des actifs (très souvent le nombre de sauts de plus d'un tick est loin d'être négligeable). De plus, le comportement du prix de transaction est lié non seulement au prix efficient, mais aussi à d'autres grandeurs, comme par exemple, le carnet d'ordres.

Quelques cas particuliers de notre modèle sont présentés dans la littérature. Le cas  $\eta = 0$ ,  $L_i = 1$ ,  $t_i = \tau_i$  pour tout  $i$  correspond au cas de l'arrondi à la partie entière qui est irréaliste à

cause d'un nombre infini d'oscillations du prix. Le cas  $\eta = 1/2$ ,  $L_i = 1$ ,  $t_i = \tau_i$  correspond au modèle étudié dans [Fu09]. Notons cependant que cette spécification de  $\eta$  et des  $L_i$  ne semble pas appropriée pour les données. Le cas  $\eta < 1/2$ ,  $L_i = 1$ ,  $t_i = \tau_i$  est mentionné dans [La07].

Par souci de clarté, nous avons imposé que le niveau des barrières à atteindre pour occasionner un changement de prix était défini seulement par  $L_i$  (voir la figure 4.1). Cependant, il aurait été possible d'envisager un modèle avec des variables  $L_i^+$  et  $L_i^-$  pour la barrière vers le haut et pour la barrière vers le bas.

Comme nous l'avons expliqué dans la section précédente,  $\eta$  quantifie d'une certaine façon l'aversion au changement de prix des participants au marché. En effet,  $\eta$  contrôle la taille des zones d'incertitude (en pourcentage de tick). Plus  $\eta$  est grand, plus le prix efficient doit être loin de l'ancien prix de transaction pour occasionner un changement de prix. Dans un certain sens, un petit  $\eta$  ( $< 1/2$ ) signifie que la taille du tick apparaît trop grande aux participants de marché et un grand  $\eta$  ( $> 1/2$ ) signifie que la taille du tick apparaît trop petite aux participants de marché.

Il existe plusieurs façons de considérer  $\eta$ , en particulier avec un point de vue de praticien. Par exemple, on peut considérer qu'aux très hautes fréquences, le carnet d'ordre ne peut "suivre" le prix efficient et il freine dans un certain sens les changements de prix. Une autre possibilité est de voir  $\eta$  comme une mesure de la profondeur en prix explorée par les volumes de transaction.

Bien que la majorité des transactions ne conduisent pas à un changement de prix (voir par exemple [GO08], [PY06]), nous avons choisi de modéliser uniquement les changements de prix et de considérer par la suite (estimation de la volatilité réalisée, évaluation de l'erreur de couverture) de conserver un schéma d'échantillonnage lié aux variations de prix (tick time sampling). [GO08] montre empiriquement et théoriquement que ce schéma est généralement préférable à un échantillonnage dans le temps des transactions ou plus communément dans le temps calendaire lorsque que l'on estime la volatilité réalisée. De plus, la question du choix de la fréquence d'échantillonnage et du prix à considérer (prix de transaction, prix de milieu de fourchette,...) n'a pas besoin d'être soulevée dans notre contexte.

Nous considérons un modèle structurel pour le bruit de microstructure et ainsi nous modélisons le prix de transaction directement à partir du prix efficient. De plus, si l'on est capable d'estimer  $\eta$ , on peut alors retrouver les vraies valeurs du prix efficient aux instants  $\tau_i$ . Ceci est très pratique pour construire des procédures d'estimation (voir la section 4.3).

## 4.3 Estimation de la volatilité et de la covolatilité intégrées

Pour  $T$  fixé, notre objectif est d'estimer la volatilité et la covolatilité intégrées sur tout intervalle de la forme  $[0, t]$ ,  $t \leq T$ . Dans notre modèle, nous travaillons avec des temps d'arrêt dont la structure dépend du prix efficient, mais aussi de variables explicatives complémentaires. Ce schéma diffère de ceux habituellement utilisés qui considèrent des suites de partitions déterministes de l'intervalle  $[0, t]$  (voir par exemple [GJ94]). Notre asymptotique considère que la taille du tick tend vers zéro. Même si cette taille est fixe sur les marchés, cette hypothèse constitue une façon de faire tendre le nombre d'observations vers l'infini. Ce type d'asymptotique a été considéré par exemple dans [DJ97], [Fu09] ou [Ro07b].

### 4.3.1 Estimation de la volatilité intégrée

La volatilité intégrée de  $(X_t)_{t \geq 0}$  sur  $[0, t]$ ,  $t \leq T$ , est définie par

$$IV_t = \int_0^t \sigma_s^2 ds.$$



Dans notre modèle, un estimateur naturel est donné par la quantité suivante :

$$RV_{\alpha,t} = \sum_{\tau_i \leq t} \left( \frac{X_{\tau_i} - X_{\tau_{i-1}}}{X_{\tau_{i-1}}} \right)^2.$$

Les  $X_{\tau_i}$  ne sont pas observés, mais peuvent être obtenus par l'intermédiaire de l'équation (4.3) modulo la connaissance de  $\eta$ . Donc la première étape consiste à estimer ce paramètre, et la seconde à le remplacer par son estimateur dans l'expression de  $RV_{\alpha,t}$ .

Nous définissons l'estimateur de  $\eta$  par

$$\hat{\eta}_{\alpha,t} = \left( 0 \vee \sum_{k=1}^m \lambda_{\alpha,t,k} u_{\alpha,t,k} \right) \wedge 1,$$

avec

$$\lambda_{\alpha,t,k} = \frac{N_{\alpha,t,k}^{(a)} + N_{\alpha,t,k}^{(c)}}{\sum_{j=1}^m [N_{\alpha,t,j}^{(a)} + N_{\alpha,t,j}^{(c)}]} \text{ et } u_{\alpha,t,k} = \frac{1}{2} \left( k \left( \frac{N_{\alpha,t,k}^{(c)}}{N_{\alpha,t,k}^{(a)}} - 1 \right) + 1 \right).$$

L'idée cachée derrière cet estimateur est la suivante: les  $u_{\alpha,t,k}$  sont des estimateurs convergents de  $\eta$  pour chaque  $k$  et les  $\lambda_{\alpha,t,k}$  sont des poids naturels. Remarquons en particulier que  $N_{\alpha,t,1}^{(c)}/N_{\alpha,t,1}^{(a)}$  est un estimateur de  $2\eta$ . Par conséquent, si  $\eta$  est plus petit que  $1/2$ , il faut s'attendre à observer plus d'alternances que de continuations dans le prix de transaction et inversement dans le cas où  $\eta$  est plus grand que  $1/2$ .

Ensuite nous définissons notre estimateur de la volatilité intégrée par

$$\widehat{RV}_{\alpha,t} = \sum_{t_i \leq t} \left( \frac{\hat{X}_{\tau_i}^t - \hat{X}_{\tau_{i-1}}^t}{\hat{X}_{\tau_{i-1}}^t} \right)^2,$$

où pour  $t_i < t$ ,

$$\hat{X}_{\tau_i}^t = P_{t_i} - \alpha \left( \frac{1}{2} - \hat{\eta}_{\alpha,t} \right) \text{sign}(P_{t_i} - P_{t_{i-1}}).$$

L'outil essentiel pour établir la convergence de nos estimateurs est le théorème de Dambis-Schwarz. Il permet de trouver un changement de temps dans lequel le processus de prix est un mouvement Brownien et permet donc d'en déduire les lois des temps de sortie des zones d'incertitude. Ensuite, nous utilisons des résultats classiques de la convergence de la variation quadratique calculée à partir de temps d'arrêt, voir par exemple le théorème I.4.47 dans [JS02].

Dans la suite,  $\xrightarrow{u.c.p.}$  désignera la convergence en probabilité sur tout compact inclus dans l'intervalle  $[0, T]$ . Nous aurons aussi recours à l'abus de notation suivant: nous dirons que  $Z^\alpha$  converge uniformément en probabilité vers  $Z$  quand  $\alpha$  tend vers zéro si pour toute séquence  $\alpha_n$  tendant vers zéro,  $Z^{\alpha_n} \xrightarrow{u.c.p.} Z$ . Nous avons le résultat suivant.

**Théorème 4.1.** *Lorsque  $\alpha$  tend vers 0, nous avons*

$$\hat{\eta}_{\alpha,t} \xrightarrow{u.c.p.} \eta \text{ et } \widehat{RV}_{\alpha,t} \xrightarrow{u.c.p.} \int_0^t \sigma_s^2 ds.$$

Pour notre prochain théorème, nous supposons que le processus  $\chi$  (qui intervient dans la distribution des  $L_i$ ) est une semi-martingale d'Itô continue  $\mathcal{F}_t$ -adaptée avec un processus de tendance prévisible et dont les trajectoires sont localement bornées, et avec une matrice de volatilité positive dont les éléments ont des trajectoires càdlàg.

Donnons maintenant la définition de la convergence stable en loi. Soit  $Z^\alpha$  une famille de variables aléatoires (prenant leurs valeurs dans un espace de fonctions càdlàg équipé de la topologie de Skorokhod  $J_1$ ). Soit  $\alpha_n$  une suite déterministe tendant vers 0 lorsque  $n$  tend vers l'infini et  $\mathcal{I}$  une

sous sigma-algèbre de  $\mathcal{F}$ . Nous disons que  $Z^{\alpha_n}$  converge en loi  $\mathcal{I}$ -stablement vers  $Z$  quand  $\alpha_n$  tend vers zéro ( $Z^{\alpha_n} \xrightarrow{\mathcal{I}-\mathcal{L}^s} Z$ ) si pour toute variable aléatoire réelle bornée  $\mathcal{I}$ -mesurable  $V$ ,  $(V, Z^{\alpha_n})$  converge en loi vers  $(V, Z)$  lorsque  $n$  tend vers l'infini. Il s'agit d'un mode de convergence plus fort que la convergence en loi (voir [JS02] pour des précisions complémentaires et des définitions équivalentes). Finalement, nous disons que  $Z^\alpha$  converge en loi  $\mathcal{I}$ -stablement vers  $Z$  lorsque  $\alpha$  tend vers zéro si pour toute suite  $\alpha_n$  tendant vers zéro,  $Z^{\alpha_n} \xrightarrow{\mathcal{I}-\mathcal{L}^s} Z$ .

Ensuite, nous introduisons les notations suivantes :  $\nabla_1$  et  $\nabla_2$  sont deux vecteurs de dimension  $(2m+1)$  définis par  $\nabla_{1,1} = 1$ ,  $\nabla_{1,i} = 0$  pour  $i = 2, \dots, 2m+1$ ,  $\nabla_{2,1} = 0$  et pour  $i = 1, \dots, m$ ,

$$\nabla_{2,2i} = i + \eta - 1/2, \quad \nabla_{2,2i+1} = -i^{-1}(i + \eta - 1/2)(i + 2\eta - 1).$$

Les processus  $(f_t)_{t \geq 0}$  et  $(\mu_t)_{t \geq 0}$  sont définis par

$$f_t = \int_0^t \varphi(\chi_u) \sigma_u^2 X_u^2 du, \quad \mu_t = \int_0^t \sum_{k=1}^m \frac{2k(k-1+2\eta)}{2k-1+2\eta} p_k(\chi_u) \varphi(\chi_u) \sigma_u^2 du,$$

avec

$$\varphi(\chi_u) = \left( \sum_{j=1}^m p_j(\chi_u) j(j-1+2\eta) \right)^{-1}.$$

Nous sommes maintenant en mesure d'énoncer notre théorème. L'idée de la preuve est de considérer un second changement de temps avant d'utiliser celui mentionné précédemment. Dans ce temps modifié, les changements de prix apparaissent asymptotiquement comme équi-répartis dans le temps. Ainsi nous pouvons utiliser ensuite des théorèmes plus classiques se basant sur des schémas d'échantillonnage déterministes. Soit  $\mathbb{D}[0, T]$  l'espace des fonctions càdlàg sur  $[0, T]$  et  $^\top$  l'opérateur de transposition. Nous avons le résultat suivant.

**Théorème 4.2.** *Soit  $\mathcal{I}$  la filtration générée par les processus  $X$  et  $\chi$ . Lorsque  $\alpha$  tend vers 0, nous avons*

$$\alpha^{-1}(\widehat{RV}_{\alpha,t} - \int_0^t \sigma_s^2 ds) \xrightarrow{\mathcal{I}-\mathcal{L}^s} (\nabla_1^\top + \frac{\mu_t}{f_t} \nabla_2^\top) \int_0^t b_{f_s} dW_{f_s},$$

dans  $\mathbb{D}[0, T]$ , où  $W$  est  $(2m+1)$  mouvement Brownien qui est défini sur une extension de l'espace de probabilité  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  et est indépendant de toutes les quantités précédentes.  $b_s$  est une matrice de taille  $(2m+1) \times (2m+1)$  présentée dans le lemme 14 du papier [A12].

### 4.3.2 Estimation de la covolatilité intégrée

Nous passons maintenant au problème de l'estimation de la covolatilité intégrée lorsque deux actifs sont observés. Dans notre contexte, il n'est pas possible d'utiliser l'estimateur usuel, la covariation réalisée pour deux raisons: les données sont asynchrones et il y a la présence du bruit de microstructure. Le problème de l'asynchronicité doit être traité avec beaucoup d'attention car l'idée d'échantillonner en utilisant le dernier prix traité conduit à un biais systématique. Cet effet est connu sous la dénomination d'effet de Epps (voir [HY05] et [Z06a] pour des détails). Cette difficulté a été traitée dans [HY05] dans le cas où les fonctions de corrélation et de volatilité sont déterministes et les dates de transactions sont indépendantes du prix efficient. Cependant, l'estimateur proposé ne semble pas robuste à la présence de bruit de microstructure (voir par exemple [VL07]). Nous montrons qu'une modification de l'estimateur de Hayashi-Yoshida qui utilise les valeurs estimées données par l'équation (4.3) est convergent.

Plus précisément, nous considérons un processus bidimensionnel  $(X_t^{(1)}, X_t^{(2)})$  qui est une  $\mathcal{F}_t$ -adaptée semi-martingale d'Itô continue telle que pour  $j = 1, 2$

$$Y_t^{(j)} = \log X_t^{(j)} = \log X_0^{(j)} + \int_0^t a_u^{(j)} du + \int_0^t \sigma_u^{(j)} dW_u^{(j)}$$

et

$$\langle W^{(1)}, W^{(2)} \rangle_t = \int_0^t \rho_s ds,$$

où  $\rho_s$  est un processus adapté avec des trajectoires càdlàg tel que, pour tout  $s$ ,  $-1 < \rho_s < 1$ . Nous imposons les mêmes conditions sur  $a^{(j)}$  et  $\sigma^{(j)}$  que celles imposées dans la section 4.2.1. Les quantités  $\alpha^{(j)}$ ,  $\eta^{(j)}$ ,  $L_i^{(j)}$ ,  $\tau_i^{(j)}$ ,  $t_i^{(j)}$  et  $P_{t_i^{(j)}}^{(j)}$  sont aussi définies de la même façon que dans la section 4.2.1.

L'estimateur usuel de la covariation de Hayashi-Yoshida est donné par

$$HY_t = \sum_{t_{i_1}^{(1)} \leq t} \sum_{t_{i_2}^{(2)} \leq t} (\log(P_{t_{i_1}^{(1)}}^{(1)}) - \log(P_{t_{i_1-1}^{(1)}}^{(1)})) (\log(P_{t_{i_2}^{(2)}}^{(2)}) - \log(P_{t_{i_2-1}^{(2)}}^{(2)})) \mathbb{I}_{[t_{i_1-1}^{(1)}, t_{i_1}^{(1)}] \cap [t_{i_2-1}^{(2)}, t_{i_2}^{(2)}] \neq \emptyset}. \quad (4.5)$$

Notre estimateur est donc défini de la manière suivante

$$\widehat{RCV}_t = \sum_{\tau_{i_1}^{(1)} \leq t} \sum_{\tau_{i_2}^{(2)} \leq t} (\log(\hat{X}_{\tau_{i_1}^{(1)}}^{(1)}) - \log(\hat{X}_{\tau_{i_1-1}^{(1)}}^{(1)})) (\log(\hat{X}_{\tau_{i_2}^{(2)}}^{(2)}) - \log(\hat{X}_{\tau_{i_2-1}^{(2)}}^{(2)})) \mathbb{I}_{[\tau_{i_1-1}^{(1)}, \tau_{i_1}^{(1)}] \cap [\tau_{i_2-1}^{(2)}, \tau_{i_2}^{(2)}] \neq \emptyset}.$$

Nous avons le résultat suivant.

**Théorème 4.3.** *Supposons que  $\alpha^{(2)} = c\alpha^{(1)}$  avec  $c > 0$ . Lorsque  $\alpha^{(1)}$  tend vers zéro, nous avons*

$$\widehat{RCV}_t \xrightarrow{u.c.p.} \int_0^t \rho_s \sigma_s^{(1)} \sigma_s^{(2)} ds.$$

Ainsi, le problème de l'estimation de la covolatilité intégrée est un autre exemple qui montre que notre méthodologie consistant à retrouver les prix efficients puis à utiliser des procédures statistiques classiques dans un cadre sans bruit de microstructure est une méthodologie pertinente. Finalement signalons qu'un théorème pour la loi limite serait certainement très compliqué à obtenir avec notre méthodologie puisqu'il nécessiterait un changement de temps adapté aux processus de prix des deux actifs.

## 4.4 Erreurs de couverture en présence de bruit de microstructure

Nous considérons un produit dérivé européen de maturité  $T$  et de fonction de paiement  $F(X_T)$ , où  $F$  est une fonction suffisamment régulière. Nous présentons dans cette section la stratégie de référence et les deux stratégies adaptées à notre marché avec zones d'incertitude. Puis nous donnons les erreurs asymptotiques de couverture. Nous supposons que les autres hypothèses d'un marché sans friction sont satisfaites. En particulier, l'actif sans risque et l'actif risqué sont parfaitement divisibles. Nous supposons sans perte de généralité que le taux sans risque est nul.

### 4.4.1 Stratégie de couverture de référence

La stratégie de référence est celle d'un agent décidant (peut être à tort) que la volatilité du prix efficient à la date  $t$  est de la forme  $\sigma(t, X_t)$ , où  $\sigma$  est une fonction suffisamment régulière. Une telle hypothèse permet lorsqu'elle est vraie de construire un portefeuille de répliation autofinçant composé de l'actif risqué et de l'actif sans risque et dont le prix à la date  $t$  est unique et de la forme  $C(t, X_t)$ . La fonction  $C$  satisfait

$$\dot{C}_t(t, x) + \frac{1}{2} \sigma^2(t, x) x^2 \ddot{C}_{xx}(t, x) = 0, \quad C(T, x) = F(x)$$

avec  $\dot{C}_t(t, x) = \partial C(t, x) / \partial t$ ,  $\ddot{C}_{xx}(t, x) = \partial^2 C(t, x) / \partial x^2$ .

Supposons que l'agent implémente cette stratégie dans un marché sans friction. Cela conduit à notre portefeuille de référence de valeur  $\Pi_t$  satisfaisant

$$\Pi_t = C(0, X_0) + \int_0^t \dot{C}_x(u, X_u) dX_u.$$

Notons que, si le modèle est mal spécifié, alors  $\Pi_t$  est différent de  $C(t, X_t)$ , voir [EJS98]. Finalement, nous supposons qu'il existe une constante positive  $K$  telle que  $C(t, x)$  satisfasse

$$\left| \frac{\partial^{\gamma+\beta} C(t, x)}{\partial t^\gamma \partial x^\beta} \right| \leq K,$$

pour tout  $(t, x) \in [0, T] \times [0, \infty)$ ,  $\gamma \in \mathbb{N}$ ,  $\beta \in \mathbb{N}$  et  $0 \leq \gamma + \beta \leq 3$ .

#### 4.4.2 Stratégies de couverture dans le modèle avec zones d'incertitude

Nous supposons qu'il est toujours possible d'acheter ou de vendre au dernier prix traité quelque soit le volume demandé et son signe. Ainsi, nous imposons naturellement que les dates où la composition du portefeuille de réplication peut être remise à jour sont les dates où les prix de transaction changent. Par simplification, nous négligeons le délai de réaction et posons  $t_i = \tau_i$ . Par conséquent, les stratégies sont constantes par morceaux. Nous imposons que la quantité d'actif risqué détenue après reconstitution du portefeuille à la date  $\tau_i$  est  $\dot{C}_x(\tau_i, X_{\tau_i})$ . Une telle stratégie est possible dès lors que  $\eta$  est connu et permet de calculer  $X_{\tau_i}$ . Ceci n'est pas restrictif dans la mesure où ce paramètre peut être estimé de manière très précise (voir la section précédente).

Nous allons considérer les deux stratégies de couverture suivantes : (i) la composition du portefeuille de couverture est remise à jour chaque fois que le prix de transaction change, (ii) la composition du portefeuille de couverture est remise à jour uniquement que lorsque le prix de transaction a varié d'au moins une certaine valeur.

#### 4.4.3 Distributions asymptotiques des erreurs de couverture

Dans notre cadre de modélisation, l'erreur de couverture est causée par :

- un nombre fini de réallocations,
- le bruit de microstructure qui perturbe le prix efficient: les variations du prix de transaction (multiple du tick de l'actif) diffèrent des variations du prix efficient.

Nous analysons cette erreur en deux temps. Tout d'abord, nous supposons qu'il n'y a pas d'erreur due au bruit de microstructure bien que les dates de transactions soient endogènes (pour tout  $i$ ,  $P_{\tau_i} = X_{\tau_i}$ ) : nous nous attendons à des résultats plus ou moins proches de [BKL00], [HM05] et [TV08] où des temps de transaction exogènes sont considérés. Dans un second temps, nous rajoutons l'erreur due au bruit de microstructure et discutons les deux stratégies implémentées.

Notre asymptotique est de considérer que la taille du tick tend vers zéro. Même si dans la pratique, la taille du tick est fixée, cela nous permet de nous placer dans un cadre mathématique où le nombre de reconstitutions du portefeuille tend vers l'infini.

#### Erreur de couverture sans bruit de microstructure

Soit  $\phi(t) = \sup\{\tau_i : \tau_i < t\}$ . En l'absence de bruit de microstructure, l'erreur de couverture est donnée par

$$L_{\alpha, t}^{(1)} = \int_0^t [\dot{C}_x(u, X_u) - \dot{C}_x(\phi(u), X_{\phi(u)})] dX_u.$$

Soit  $f_t$  la limite en probabilité de  $\alpha^2 N_{\alpha,t}$ , lorsque  $\alpha \rightarrow 0$ . Nous avons d'après la section précédente

$$f_t = \int_0^t \varphi(\chi_u) \sigma_u^2 X_u^2 du \quad \text{où} \quad \varphi(\chi_u) = \left( \sum_{j=1}^m p_j(\chi_u) j(j-1+2\eta) \right)^{-1}.$$

La réciproque de  $f$ , notée  $\theta$ , est un changement de temps pour lequel les observations sont asymptotiquement uniformément réparties sur n'importe quel intervalle. Posons  $\Delta X_{\tau_i} = X_{\tau_{i+1}} - X_{\tau_i}$  et  $(i_\alpha)$  une suite telle que  $\tau_{i_\alpha}$  tend vers  $t$  en probabilité quand  $\alpha \rightarrow 0$ . Nous introduisons alors le quatrième moment conditionnel du changement de prix normalisé par  $\alpha$  à la date  $t$  :

$$\mu_4(\chi_t) = \lim_{\alpha \rightarrow 0} \mathbb{E}_{\mathcal{E}_{\tau_{i_\alpha}}} [\alpha^{-4} (\Delta X_{\tau_{i_\alpha}})^4] = \sum_{k=1}^m k(k-1+2\eta)(k^2 - k + 2\eta k + 4\eta^2 - 4\eta + 1) p_k(\chi_t).$$

Soit  $\mathcal{I}$  la filtration générée par les processus  $X$  and  $\chi$ . Nous avons le résultat suivant.

**Théorème 4.4.** *Lorsque  $\alpha$  tend vers 0,*

$$N_{\alpha,t}^{1/2} L_{\alpha,t}^{(1)} \xrightarrow{\mathcal{I}\text{-}\mathcal{L}^s} L_t^{(1)} := f_t^{1/2} \int_0^t c_{f_s}^{(1)} dW_{f_s}^{(1)}, \quad (4.6)$$

dans  $\mathbb{D}[0,T]$ , où  $W^{(1)}$  est un mouvement Brownien défini sur une extension de l'espace de probabilité  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  et est indépendant de toutes les quantités précédentes, et  $c_s^{(1)}$  est tel que

$$(c_s^{(1)})^2 = \frac{1}{6} \ddot{C}_{xx}^2(\theta_s, X_{\theta_s}) \mu_4(\chi_{\theta_s}).$$

Nous remarquons que l'erreur de couverture est proportionnelle au nombre de remises à jour et dépend du gamma<sup>1</sup> du produit dérivé dans le modèle à volatilité locale. Notons également, que, si la composition du portefeuille de couverture est remise à jour à des dates exogènes  $\tau_i = i\alpha^2 T$ , pour  $i = 1, \dots, n_\alpha = \alpha^{-2}$ , nous aurions alors  $f_t = t$ ,  $\mu_4(X_t) = 3\sigma_t^4 X_t^4$  et la distribution de l'erreur de couverture asymptotique aurait été identique à celle de [BKL00]. Nous en concluons que, bien que nous considérons des dates endogènes de transaction, nos résultats sont sur la même ligne que ceux de [BKL00] et [HM05].

### Erreur de couverture avec bruit de microstructure

En présence de bruit de microstructure, le prix de transaction diffère du prix efficient. L'erreur de couverture est désormais donnée par

$$L_{\alpha,t}^{(2)} = \int_0^t \dot{C}_x(u, X_u) dX_u - \int_0^t \dot{C}_x(\phi(u), X_{\phi(u)}) dP_u.$$

Soit

$$\pi_a(\chi_t) = \lim_{\alpha \rightarrow 0} \mathbb{E}_{\mathcal{E}_{\tau_{i_\alpha}}} [1\{\Delta X_{\tau_{i_\alpha}} \Delta X_{\tau_{i_\alpha-1}} < 0\}] = \sum_{k=1}^m \frac{k}{2k-1+2\eta} p_k(\chi_t)$$

la probabilité asymptotique conditionnelle que le prochain changement de prix à la date  $t$  soit une alternance et soit

$$\mu_{1,a}^*(\chi_t) = \lim_{\alpha \rightarrow 0} \mathbb{E}_{\mathcal{E}_{\tau_{i_\alpha}}} [\alpha^{-1} |\Delta X_{\tau_{i_\alpha}}| 1\{\Delta X_{\tau_{i_\alpha}} \Delta X_{\tau_{i_\alpha-1}} < 0\}] = \sum_{k=1}^m \frac{k(k-1+2\eta)}{2k-1+2\eta} p_k(\chi_t)$$

---

1. Le gamma d'une option mesure la sensibilité à l'ordre 2 du prix de l'option aux variations de l'actif sous-jacent.

l'espérance conditionnelle de la valeur absolue du changement de prix normalisé à la date  $t$  lorsque ce changement de prix est une alternance. Nous avons le résultat suivant.

**Théorème 4.5.** *Lorsque  $\alpha$  tend vers 0,*

$$L_{\alpha,t}^{(2)} \xrightarrow{\mathcal{I}-\mathcal{L}^s} L_t^{(2)} := \int_0^t b_{f_s}^{(2)} dX_s + \int_0^t c_{f_s}^{(2)} d\mathcal{W}_{f_s}^{(2)},$$

dans  $\mathbb{D}[0,T]$ , où  $\mathcal{W}^{(2)}$  est un mouvement Brownien défini sur une extension de l'espace de probabilité  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  et est indépendant de toutes les quantités précédentes, et  $b_s^{(2)}$  et  $c_s^{(2)}$  sont tels que

$$\begin{aligned} b_s^{(2)} &= (1 - 2\eta) \dot{C}_x(\theta_s, X_{\theta_s}) \mu_{1,a}^*(\chi_{\theta_s}) \varphi(\chi_{\theta_s}) \\ (c_s^{(2)})^2 &= (1 - 2\eta)^2 \dot{C}_x^2(\theta_s, X_{\theta_s}) \varphi(\chi_{\theta_s}) (\pi_a(\chi_{\theta_s}) \varphi^{-1}(\chi_{\theta_s}) - (\mu_{1,a}^*(\chi_{\theta_s}))^2). \end{aligned}$$

Il est très important de noter que l'erreur de couverture avec bruit de microstructure n'est pas normalisée comme dans le cas précédent. Cela signifie que cette erreur ne disparaît pas même lorsque le nombre de recompositions du portefeuille tend vers l'infini. Par conséquent elle est du même ordre de grandeur que l'erreur due à l'utilisation d'un mauvais modèle. Cependant si  $\eta = 1/2$ , les variations du prix de transaction coïncident avec les variations du prix efficient aux moments de sortie des zones d'incertitude ( $\Delta P_{\tau_i} = \Delta X_{\tau_i}$ ) et l'erreur de couverture disparaît dans ce cas précis.

La première composante de l'erreur de couverture asymptotique,  $\int_0^t b_{f_s}^{(2)} dX_s$ , est causée par une asymétrie entre les alternances et les continuations. En effet, lorsqu'une alternance se produit,  $\Delta P_{\tau_i} - \Delta X_{\tau_i} = (1 - 2\eta) \text{sign}(\Delta X_{\tau_i})$  alors que, lorsqu'une continuation se produit  $\Delta P_{\tau_i} - \Delta X_{\tau_i} = 0$ . De plus, notons que la variation quadratique de l'erreur de couverture asymptotique est égale à

$$(1 - 2\eta)^2 \int_0^t \dot{C}_x^2(s, X_s) \pi_a(\chi_s) df_s.$$

Elle dépend maintenant du delta<sup>2</sup> du produit dérivé dans le modèle à volatilité locale et de la proportion d'alternances. Par conséquent, la variance de l'erreur de couverture en présence de bruit de microstructure augmente avec la position en delta. Contrairement au cas précédent, on se retrouve maintenant avec une erreur du premier ordre. Finalement, notons que cette erreur n'est pas nécessairement centrée à cause de la présence de  $\int_0^t b_{f_s}^{(2)} dX_s$  dans la limite.

## Fréquences de réallocations optimales en présence de bruit de microstructure

Nous construisons maintenant une stratégie où le portefeuille de couverture est remis à jour uniquement lorsque le prix a suffisamment évolué<sup>3</sup>. Il s'agit d'une stratégie peut être plus naturelle que celle basée sur toutes les remises à jour car, d'après les résultats précédents, cette dernière stratégie conduit à supporter une erreur du premier ordre. C'est pourquoi, à partir de maintenant, nous supposons que le portefeuille n'est remis à jour que lorsque le prix change d'au moins  $l_\alpha$  ticks. Par simplicité, nous supposerons ici que  $m = 1$ . Nous allons choisir  $l_\alpha$  tel que  $l_\alpha \rightarrow \infty$  et  $\alpha l_\alpha \rightarrow 0$  quand  $\alpha \rightarrow 0$ . De cette façon, l'erreur de couverture due au bruit de microstructure sur le prix et l'erreur de couverture due à nombre fini de réallocations vont disparaître toutes les deux. Remarquons que la stratégie est plus ou moins équivalente à se placer sur une sous-grille de prix de taille de tick  $\alpha l_\alpha$  et d'attendre pour remettre à jour le portefeuille que le prix ait varié de ce nouveau tick.

2. Le delta de l'option mesure la sensibilité du prix de l'option par rapport au cours de l'actif sous-jacent.

3. Notons que nous pourrions également envisager une recomposition du portefeuille que lorsque le delta du produit a suffisamment évolué.

Soit  $N_{\alpha,t}^{(l)}$  le nombre de recompositions du portefeuille entre 0 et  $t$ . Nous définissons récursivement la suite  $(\tau_i^{(l)})_{i \geq 1}$  par

$$\tau_{i+1}^{(l)} = \inf \left\{ t : t > \tau_i^{(l)}, X_t = X_{\tau_i^{(l)}} - \alpha(l_\alpha - \frac{1}{2} + \eta) \text{ ou } X_t = X_{\tau_i^{(l)}} + \alpha(l_\alpha - \frac{1}{2} + \eta) \right\}.$$

Soit  $f_t^{(l)}$  la limite en probabilité de  $(\alpha l_\alpha)^2 N_{\alpha,t}^{(l)}$  lorsque  $\alpha \rightarrow 0$ . Nous avons

$$f_t^{(l)} = \int_0^t \sigma_u^2 X_u^2 du.$$

Soit  $\phi^{(l)}(t) = \sup\{\tau_i^{(l)} : \tau_i^{(l)} < t\}$ . L'erreur de couverture est maintenant donnée par la formule :

$$L_{\alpha,t}^{(3)} = \int_0^t \dot{C}_x(u, X_u) dX_u - \int_0^t \dot{C}_x(\phi^{(l)}(u), X_{\phi^{(l)}(u)}) dP_u.$$

Soit  $\Delta X_{\tau_{i-1}^{(l)}} = X_{\tau_i^{(l)}} - X_{\tau_{i-1}^{(l)}}$ . Cette erreur peut être décomposée de la manière suivante :

$$\begin{aligned} L_{\alpha,t}^{(3)} &= L_{\alpha,t}^{(3,1)} + L_{\alpha,t}^{(3,2)} + L_{\alpha,t}^{(3,3)} := \sum_{i=1}^{N_{\alpha,t}^{(l)}} \int_{\tau_{i-1}^{(l)}}^{\tau_i^{(l)}} (\dot{C}_x(u, X_u) - \dot{C}_x(\tau_{i-1}^{(l)}, X_{\tau_{i-1}^{(l)}})) dX_u \\ &\quad + \sum_{i=1}^{N_{\alpha,t}^{(l)}} \dot{C}_x(\tau_{i-1}^{(l)}, X_{\tau_{i-1}^{(l)}}) (\Delta X_{\tau_{i-1}^{(l)}} - \Delta P_{\tau_{i-1}^{(l)}}) \\ &\quad + \int_{\tau_{N_{\alpha,t}^{(l)}}^{(l)}}^t \dot{C}_x(u, X_u) dX_u - \dot{C}_x(\tau_{N_{\alpha,t}^{(l)}}^{(l)}, X_{\tau_{N_{\alpha,t}^{(l)}}^{(l)}}) (P_t - P_{\tau_{N_{\alpha,t}^{(l)}}^{(l)}}). \end{aligned}$$

Nous avons le résultat suivant.

**Théorème 4.6.** *Lorsque  $\alpha$  tend vers 0,*

$$\begin{aligned} (\alpha l_\alpha)^{-1} L_{\alpha,t}^{(3,1)} &\xrightarrow{\mathcal{I}-\mathcal{L}^s} L_t^{(3,1)} := \int_0^t b_{f_s^{(l)}}^{(3,1)} dX_s + \int_0^t c_{f_s^{(l)}}^{(3,1)} dW_{f_s^{(l)}}^{(3,1)}, \\ l_\alpha L_{\alpha,t}^{(3,2)} &\xrightarrow{\mathcal{I}-\mathcal{L}^s} L_t^{(3,2)} := \int_0^t b_{f_s^{(l)}}^{(3,2)} dX_s + \int_0^t c_{f_s^{(l)}}^{(3,2)} dW_{f_s^{(l)}}^{(3,2)}, \\ (l_\alpha \vee (\alpha l_\alpha)^{-1}) L_{\alpha,t}^{(3,3)} &\xrightarrow{\mathcal{I}-\mathcal{L}^s} 0, \end{aligned}$$

dans  $\mathbb{D}[0, T]$ , où  $\mathcal{W}_{f_s^{(l)}}^{(3,1)}$  et  $\mathcal{W}_{f_s^{(l)}}^{(3,2)}$  sont des mouvements Browniens définis sur une extension de l'espace de probabilité  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  et sont indépendants de toutes les quantités précédentes, et  $b_s^{(3,1)}$  et  $c_s^{(3,1)}$  sont tels que

$$b_s^{(3,1)} = 0, \quad (c_s^{(3,1)})^2 = \frac{1}{6} \ddot{C}_{xx}^2(\theta_s, X_{\theta_s}),$$

et  $b_s^{(3,2)}$  et  $c_s^{(3,2)}$  sont tels que

$$b_s^{(3,2)} = \frac{(1-2\eta)}{2} \dot{C}_x(\theta_s, X_{\theta_s}), \quad (c_s^{(3,2)})^2 = \frac{(1-2\eta)^2}{4} \dot{C}_x^2(\theta_s, X_{\theta_s}).$$

Le choix optimal de  $l_\alpha$  résulte alors d'un ‘‘équilibrage’’ entre les variances des erreurs asymptotiques renormalisées.

**Théorème 4.7.** Soit  $l_\alpha = \alpha^{-1/2}$ . Lorsque  $\alpha$  tend vers 0, nous avons

$$(N_{\alpha,t}^{(l)})^{1/4} L_{\alpha,t}^{(3)} \xrightarrow{\mathcal{I}-\mathcal{L}^s} L_t^{(3)} := (f_t^{(l)})^{1/4} \left( \int_0^t b_{f_s^{(l)}}^{(3)} dX_s + \int_0^t c_{f_s^{(l)}}^{(3)} d\mathcal{W}_{f_s^{(l)}}^{(3)} \right)$$

dans  $\mathbb{D}[0,T]$ , où  $\mathcal{W}^{(3)}$  est un mouvement Brownien défini sur une extension de l'espace de probabilité  $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  et est indépendant de toutes les quantités précédentes, et  $b_s^{(3)}$  et  $c_s^{(3)}$  sont tels que

$$\begin{aligned} b_s^{(3)} &= \frac{(1-2\eta)}{2} \dot{C}_x(\theta_s, X_{\theta_s}) \\ (c_s^{(3)})^2 &= \frac{(1-2\eta)^2}{4} \dot{C}_x^2(\theta_s, X_{\theta_s}) + \frac{1}{6} \ddot{C}_{xx}^2(\theta_s, X_{\theta_s}). \end{aligned}$$

Cette stratégie optimale permet de réduire significativement l'erreur de couverture en présence de bruit de microstructure. La variance asymptotique de cette erreur dépend à la fois du delta et du gamma du produit dérivé dans le modèle à volatilité locale.

Remarquons que le choix optimal de  $l_\alpha$  est du même ordre de grandeur que la racine carrée du nombre de fois où la composition du portefeuille de couverture est remise à jour, et est du même ordre de grandeur que l'inverse de la variance asymptotique de l'erreur de couverture. Asymptotiquement, pour décroître la variance d'un facteur 2, on doit realouer les actifs du portefeuille 4 fois plus souvent avec un niveau cible ( $\alpha l_\alpha$ ) divisé par 2.

## 4.5 Autres travaux et projets de recherche

Dans [A15], nous considérons deux processus Gaussiens à accroissements indépendants intimement reliés : l'un est corrélé partiellement à une version retardée de l'autre. Cette situation apparaît naturellement en finance lorsque deux actifs partagent les mêmes fondamentaux économiques, mais que leur prix ne réagissent pas forcément en même temps aux nouvelles informations. Généralement, l'un réagit avant l'autre, conduisant à une relation de "lead-lag" entre les deux (un meneur, un suiveur). De nombreuses raisons peuvent expliquer ce décalage dans le temps entre les deux actifs: des mécanismes de transactions différents, des intervenants différents, des coûts de transactions différents, des asymétries d'information... En pratique, cette relation de retard peut être considérée comme une inefficience du marché, et certains agents peuvent être intéressés pour estimer le retard et profiter d'opportunités d'arbitrage.

Le problème de l'estimation du paramètre de décalage a été peu étudié dans la littérature. Une première approche est proposée dans [HRY09]. Cette approche est basée sur l'étude de l'auto-corrélogramme croisé entre les deux processus calculé sur une journée. Dans [A15], nous privilégions une approche basée sur la Théorie des Matrices Aléatoires. Nous caractérisons en effet la distribution spectrale limite de la matrice de variances-covariances des accroissements des deux processus lorsque le pas de temps entre deux observations tend vers zéro. Nous en déduisons par le théorème de Marcenko-Pastur (voir [MP67]) la distribution spectrale limite de la matrice empirique de variances-covariances. Le résultat est obtenu lorsque le nombre de trajectoires considéré est proportionnel au nombre d'accroissements. Finalement, on utilise le moment d'ordre 2 de la distribution limite pour construire un estimateur du paramètre de décalage et montrons que cet estimateur est asymptotiquement Gaussien. Nous projetons d'étudier bientôt des modèles plus réalistes de décalages temporels endogènes.

Nous souhaitons enfin travailler dans l'avenir sur les processus multi-fractals et leurs applications en finance. Plus précisément, nous pensons étudier les propriétés des variations et des extrêmes de ces processus à différentes échelles.



# Bibliographie

- [ANT00] ANCONA-NAVARRETE, M., TAWN, J.A. (2000). A comparison of methods for estimating the extremal index. *Extremes*, 3, 5-38.
- [As00] ASMUSSEN, S. (2000). *Ruin Probabilities*. World Scientific.
- [As87] ASMUSSEN, S. (1987). *Applied Probability and Queues*. Springer.
- [Ba88] BALL, C. (1988). Estimation bias induced by discrete security prices. *Journal of Finance*, 43, 841-865.
- [BKR99] BALKEMA, A.A., KLÜPPELBERG, C., RESNICK, S.I. (1999). Limit laws for exponential families. *Bernoulli*, 1, 951-958.
- [BKR03] BALKEMA, A.A., KLÜPPELBERG, C., RESNICK, S.I. (2003). Domains of attraction for exponential families and asymptotic behavior of Laplace transforms. *Stochastic Processes and their Applications*, 107, 83-103.
- [BSh02] BARNDORFF-NIELSEN, O. E., SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, 64, 253-280.
- [BSh07] BARNDORFF-NIELSEN, O. E., SHEPHARD, N. (2007). Variation, jumps, market frictions and high frequency data in financial econometrics. *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*, (edited by Richard Blundell, Persson Torsten and Whitney K Newey), Econometric Society Monographs, Cambridge University Press.
- [BHLS09] BARNDORFF-NIELSEN O., HANSEN P., LUNDE A., SHEPHARD, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76, 1481-1536.
- [BPP84] BEARD, R., PENTIKAINEN, R., PESONEN, E. (1984). *Risk Theory*. Chapman and Hall.
- [BGST04] BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J., TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.
- [BKL00] BERTSIMAS, D., KOGAN, L., LO, A.W. (2000). When is time continuous?. *Journal of Financial Economics*, 55, 173-204.
- [BK01] BORKOVEC, M., KLÜPPELBERG, C. (2001). The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Annals of Applied Probability*, 11, 1220-1241.
- [BPSZ06] BOXMA, O., PERRY, D., STADJE, W., ZACKS, S. (2006). A Markovian growth-collapse model. *Advances in Applied Probabilities*, 38, 221-243.
- [BG03] BUCHMANN, B. GRÜBEL, R. (2003). Decompounding: an estimation problem for Poisson random sums. *Annals of Statistics*, 31, 1054-1074.
- [BG04] BUCHMANN, B. GRÜBEL, R. (2004). Decompounding Poisson random sums: recursively truncated estimates in the discrete case. *Annals of the Institute of Statistical Mathematics*, 56, 743-756.
- [Bu96] BÜHLMANN, H. (1996). *Mathematical Methods of Risk Theory*. Springer, Berlin.
- [CG00] CAI, J., GARRIDO, J. (2000). Two-sided bounds for tails of compound negative binomial distributions in the exponential and heavy-tailed cases. *Scandinavian Actuarial Journal*, 100, 102-120.
- [CST09] CETIN, U., SONER, M., TOUZI, N. (2009). Option hedging under liquidity costs. To appear in *Finance and Stochastics*.
- [CF88] CHO, D., FREES, E. (1988). Estimating the volatility of discrete stock prices. *Journal of Finance*, 43, 451-466.
- [DJ97] DELATTRE, S., JACOD, J. (1997). A central limit theorem for normalized functions of the increments of a diffusion process in the presence of round-off errors. *Bernoulli*, 3, 1-28.
- [DG94] DIEBOLT, J., GUEGUAN, D. (1994). Probabilistic properties of the Beta-ARCH model. *Statistica Sinica*, 2, 157-174.
- [DFMS04] DOUC, R., FORT, G., MOULINES, E., SOULIER, P. (2004). Practical drift conditions for subgeometric rates of convergence. *Annals of Applied Probability*, 14, 1353-1377.
- [Dr03] DREES, H. (2003). Extreme quantile estimation for dependent data with applications to finance. *Bernoulli*, 9, 617-657.

- [EJS98] EL KAROUI, N., JEANBLANC-PICQUÉ, M., SHREVE, S. (1998). Robustness of the Black and Scholes formula. *Mathematical Finance*, 8, 93-126.
- [EK03] ELIAZAR, I., KLAFTER, J. (2003). A growth-collapse model: Lévy inflow, geometric crashes, and generalized Ornstein-Uhlenbeck dynamics. *Physica A*, 334, 1-21.
- [EK05] ELIAZAR, I., KLAFTER, J. (2005). Stochastic Ornstein-Uhlenbeck capacitors. *Journal of Statistical Physics*, 118, 177-198.
- [EJMT85] EMBRECHTS, P., JENSEN, J.L., MAEJIMA, M., TEUGELS, J.L. (1985). Approximations for compound Poisson and Polya processes. *Advances in Applied Probabilities*, 17, 623-637.
- [EKM97] EMBRECHTS, P., KLÜPPELBERG, C., MIKOSCH, T. (1997). *Modelling Extremal Events*. Springer, Applications of Mathematics.
- [EMT85] EMBRECHTS, P., MAEJIMA, M., TEUGELS, J.L. (1985). Asymptotic behavior of compound distributions. *ASTIN Bulletin*, 15, 45-48.
- [EV82] EMBRECHTS, P., VERAVERBEKE, N. (1982). Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance Mathematics and Economics*, 1, 55-72.
- [EV88] EMBRECHTS, P., VILLASENOR, J. (1988). Ruin estimates for large claims. *Insurance Mathematics and Economics*, 7, 269-274.
- [En00] ENGLE, R. (2000). The econometrics of ultra high frequency data. *Econometrica*, 68, 1-22.
- [Fe03] FERRO, C.A.T. (2003). *Statistical Methods for Clusters of Extreme Values*. Phd Thesis, Lancaster University.
- [FS03] FERRO, C.A.T., SEGERS, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society B*, 65, 545-556.
- [FK00] FOSS, S., KORSHUNOV, D. (2000). Sampling at a random time with a heavy-tailed distribution. *Markov Processes and Related Fields*, 6 (2000), 643-658.
- [FZ03] FOSS, S., ZACHARY, S. (2003). The maximum on a random time interval of a random walk with long-tailed increments and negative drift. *Annals of Applied Probability*, 13, 37-53.
- [Fu09] FUKASAWA, M. (2007). Central limit theorem for the realized volatility based on tick time sampling. A paraître dans *Finance and Stochastic*.
- [Ga87] GALAMBOS, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. 2nd ed. Melbourne: Krieger.
- [GJ94] GENON-CATALOT, V., JACOD, J. (1994). Estimation of the diffusion coefficient for diffusion processes: random sampling. *Scandinavian Journal of Statistics*, 21, 193-221.
- [Go89] GOLDIE, C.M. (1989). Implicit renewal theory and tails of solutions of random equations. *Annals of Applied Probabilities*, 1, 126-166.
- [GJ01] GOURIEROUX, C., JASIAK, J. (2001). *Financial Econometrics: Problems, Models, and Methods*. Princeton University Press.
- [GO08] GRIFFIN, J.E., OOMEN, R.C.A (2008). Sampling returns for realized variance calculations: tick time or transaction time. A paraître dans *Econometric Reviews*.
- [HRRV89] HAAN, L. DE, RESNICK, S., ROOTZEN, H., VRIES, C.G. DE (1989). Extremal behaviour of solutions to a stochastic difference equation with application to ARCH processes. *Stochastic Processes and their Applications*, 32, 213-224.
- [HL06] HANSEN, P.R., LUNDE, A. (2006). Realized variance and market microstructure noise. *Journal of Business and Economics Statistics*, 24, 127-161.
- [HM05] HAYASHI, T., MYKLAND, P.A. (2005). Hedging errors: an asymptotic approach. *Mathematical Finance*, 15, 309-343.
- [HY05] HAYASHI, T., YOSHIDA, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11, 359-379.
- [HRY09] HOFFMANN, M., ROSENBAUM, M., YOSHIDA, N. (2009). Lead-lag detection from non-synchronous data. Working paper.
- [HHS07] HSIEH, M., HURVICH, C.M., SOULIER, P. (2007). Asymptotic properties for duration driven long range dependent processes. *Journal of Econometrics*, 141, 913-949.
- [Hs89] HSING, T.. (1989). Extreme value theory for multivariate stationary sequences. *Journal of Multivariate Analysis*, 29, 274-291.
- [Hs91] HSING, T. (1991). Estimating the parameters of rare events. *Stochastic Processes and their Applications*, 37, 117-139.
- [Hs93a] HSING, T. (1993a). Extremal index estimation for a weakly dependent stationary sequence. *Annals of Statistics*, 21, 2043-2071.
- [Hs93b] HSING, T. (1993b). On some estimates based on sample behavior near high level excursions. *Probability Theory and Related Fields*, 95, 331-356.
- [HHL88] HSING, T., HÜSLER, J., LEADBETTER, M.R. (1988). On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields*, 78, 97-112.

- [HKK05] HSING, T., KLÜPPELBERG, C., KUHN, G. (2005). Dependence estimation and visualization in multivariate extremes with applications to financial data. *Extremes*, 7, 99-121.
- [Ja06] JACOD, J. (2006). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Processes and their Applications*, 118, 517-559.
- [JS02] JACOD, J., SHIRYAEV, A.N. (2002). *Limit Theorems for Stochastic Processes*. Second Edition, Springer-Verlag.
- [JLMPV09] JACOD, J., LI, Y., MYKLAND, P.A., PODOLSKIJ, M., VETTER, M. (2009). Microstructure noise in the continuous case: the pre-averaging Approach. *Stochastic Processes and their Applications*, 119, 2249-2276.
- [Je95] JENSEN, J.L. (1995). *Saddlepoint Approximations*. Oxford University Press.
- [Jo97] JOE, H. (1997). *Multivariate Models and Dependence Concepts*. Monographs on Statistics and Applied Probability 73, Chapman & Hall.
- [JK95] JOUINI, E., KALLAL, H. (1995). Arbitrage in securities markets with short sales constraints. *Mathematical Finance*, 5, 197-232.
- [KS97] KABANOV, Y.M., SAFARIAN, M.M. (1997). On Leland's strategy of option pricing with transactions costs. *Finance and Stochastics*, 1, 239-250.
- [KL08] KALNINA, I., LINTON, O.B. (2008). Estimating quadratic variation consistently in the presence of correlated measurement error. *Journal of Econometrics*, 147, 47-59.
- [Ke73] KESTEN, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131, 207-248.
- [KPT99] KOEHL, P.F., PHAM, H., TOUZI, N. (1999). Hedging in discrete time under transaction costs and continuous-time limit. *Journal of Applied Probability*, 36, 163-178.
- [La07] LARGE, J. (2007). Estimating quadratic variation when quoted prices change by a constant increment. A paraître dans *Journal of Econometrics*.
- [Le83] LEADBETTER, M.R. (1983). Extremes and local dependence in stationary sequences. *Probability Theory and Related Fields*, 65, 291-306.
- [LLR83] LEADBETTER, M.R., LINDGREN, G., ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.
- [LM03] LINDSKOG, F., MCNEIL, A.J. (2003). Common poisson shock models: applications to insurance and credit risk modelling. *ASTIN Bulletin*, 33, 209-238.
- [LM07] LI, Y., MYKLAND, P.A. (2007). Are volatility estimators robust with respect to modeling assumptions?. *Bernoulli*, 13, 601-622.
- [Li00] LIU, M. (2000). Modeling long memory in stock market volatility. *Journal of Econometrics*, 99, 139-171.
- [MP67] MARCENKO, V. A., PASTUR, L.A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72, 507-536.
- [MF05a] MARTINS, A.P., FERREIRA H. (2005). The multivariate extremal index and the dependence structure of a multivariate extreme value distribution. *Test*, 14, 433-448.
- [MF05b] MARTINS, A.P., FERREIRA H. (2005). Measuring the extremal dependence. *Statistics and Probability Letters*, 73, 99-103.
- [MT93] MEYN, S.P., TWEEDIE, R.L. (1993). *Markov Chains and Stochastic Stability*. Springer, New-York.
- [Na94] NANDAGOPALAN, S. (1994). On the multivariate extremal index. *J. Res. Nat. Inst. Stand. Technol.*, 99, 543-550.
- [Ob87] O'BRIEN, G.L. (1987). Extreme values for stationary and Markov sequences. *Annals of Probability*, 15, 281-289.
- [Pa81] PANJER, H.H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin*, 12, 22-26.
- [Pa99] PARKE, W.R. (1999). What is fractional integration? *Review of Economics and Statistics* 81, 632-638.
- [Pe94] PERFEKT, R. (1994). Extremal behaviour of stationary Markov chains with applications. *Annals of Applied Probabilities*, 4, 529-548.
- [Pe97] PERFEKT, R. (1997). Extreme value theory for a class of Markov chains with values in  $\mathbb{R}^d$ . *Advances in Applied Probabilities*, 29, 138-164.
- [PY06] PHILLIPS, P., YU, J. (2006). Information loss in volatility measurement with flat price trading. Working paper.
- [Pi81] PICKANDS, J. (1981). Multivariate extreme value distributions. *Bull. Int. Statist. Inst.*, 859-878.
- [Re87] RESNICK, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New-York.
- [Ro07a] ROSENBAUM, M. (2007). A new microstructure noise index. A paraître dans *Quantitative Finance*.
- [Ro07b] ROSENBAUM, M. (2009). Integrated volatility and round-off error. A paraître dans *Bernoulli*.
- [Sc99] SCHMIDLI, H. (1999). Compound sums and subexponentiality. *Bernoulli*, 5, 999-1012.

- [SS05] SCHMIDT, R., STADTMÜLLER U. (2005). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, 33, 307-335.
- [SW94] SMITH, R.L., WEISSMAN, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society: Series B*, 56, 515-528.
- [SW96] SMITH, R.L., WEISSMAN, I. (1996). Characterization and estimation of the multivariate extremal index. Tech. Rep., University of North Carolina.
- [St73] STAM, A.J. (1973) Regular variation of the tail of a subordinated distribution. *Advances in Applied Probability*, 5, 308-327.
- [Su85] SUNDT, B. (1982). Asymptotic behavior of compound distributions and Stop-loss premiums. *ASTIN Bulletin*, 13, 89-98.
- [TV08] TANKOV, P., VOLTCHKOVA, E. (2008). Asymptotic analysis of hedging errors in models with jumps. *Stochastic Processes and their Applications*, 119, 2004-2027
- [TL86] TAQQU, M.S., LEVY, J.B. (1986). Using renewal processes to generate long-range dependence and high variability. *Dependence in Probability and Statistics*, 7389. Boston, Birkhauser.
- [Te85] TEUGELS, J.L. (1985). Approximation and estimation of some compound distributions. *Insurance: Mathematics and Economics*, 4, 143-153.
- [VL07] VOEV, V., LUNDE, A. (2007). Integrated covariance estimation using high-frequency data in the presence of noise. *Journal of Financial Econometrics*, 5, 68-104.
- [WN98] WEISSMAN, I., NOVAK, S.Y. (1998). On blocks and runs estimators of the extremal index. *Journal of Statistical Planning and Inference*, 66, 281-288.
- [Z06a] ZHANG, L. (2006a). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli*, 12, 1019-1043.
- [Z06b] ZHANG, L. (2006b). Estimating covariation: Epps effect and microstructure noise. A paraître dans *Journal of Econometrics*.
- [ZMA05] ZHANG, L., MYKLAND, P.A., AÏT-SAHALIA, Y. (2005). A Tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100, 1394-1411.
- [Zh96] ZHOU, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business and Economic Statistics*, 14, 45-52.