



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-memoires-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Université de Lorraine
Faculté des Sciences et Technologies de Nancy



Master 2 Mathématiques Fondamentales et Appliquées (MFA)
Parcours Recherche

Rapport de stage

Mesure de risque : détection du régime de crise et calcul de la Value-at-Risk

Soutenu le 19 septembre 2013

Rédigé par : Khaled SALHI

Etudiant en Master 2 MFA

Encadré par : Nicolas CHAMPAGNAT
Madalina DEACONU

INRIA
INRIA

Année universitaire 2012-2013

Remerciement

Je tiens à exprimer mes sincères remerciements à tous ceux qui ont contribué à la réussite de ce stage.

Ma profonde gratitude s'adresse à mes superviseurs Madalina DEACONU et Nicolas CHAMPAGNAT pour leur encadrement et les conseils qu'ils m'ont prodigués tout au long de ce travail de recherche.

Je tiens aussi à remercier Antoine LEJAY pour ses conseils, sa disponibilité et son aide durant ce stage.

Je remercie tout particulièrement les dirigeants de la société ALPHABILITY pour leur disponibilité et leur suivi, et pour l'occasion qu'ils m'ont offert de travailler sur des données réelles issues de la bourse de Paris.

Je tiens finalement à remercier mon cher papa, ma famille et mes amis pour leur soutien du loin et de près et pour leurs encouragements.

Résumé

Notre travail vise à étudier le comportement des rendements des actifs financiers et de mesurer le risque de marché. En utilisant le modèle de Markov caché, nous classifions les données selon le critère de crise et nous détectons les régimes de crise et de non-crise. En utilisant les lois puissances pour modéliser la queue de la distribution et en prenant en compte la classification, nous mesurons la Value-at-Risk (VAR) et nous proposons un nouvel indicateur pour la gestion de portefeuille.

Mots clés : mesure de risque, Value-at-Risk, lois puissances, modèle de Markov caché, changement de régime.

Abstract

Our work aims to study the behavior of financial assets' returns and to measure market risk. Using the hidden Markov model, we classify the data according to the criteria of crisis by detecting the crisis and non-crisis regimes. By adopting the power laws to model the tail of the distribution and by taking into account this classification, we measure the Value-at-Risk (VAR) and we propose a new indicator for portfolio management.

Keywords : Risk measurement, Value-at-Risk, power laws, hidden Markov model, regime-switching.

Table des matières

1	Introduction	7
2	Contexte du stage	9
3	Modélisation par les lois puissances	11
3.1	Introduction au modèle	11
3.2	Estimateur de Hill	12
3.3	Estimation de la fonction à variations lentes et des quantiles	20
3.4	La <i>Value-at-Risk</i> et le modèle des lois puissances	22
3.5	Etude numérique : estimation des paramètres de la loi puissance, calcul de la VaR	23
3.6	Backtesting sur la VaR	25
4	Analyse de sensibilité sur le modèle des lois puissances	29
4.1	Sensibilité par rapport à la taille de la fenêtre	29
4.2	Sensibilité par rapport aux données bruitées	31
5	Modèle de Markov Caché	38
5.1	Problématique	38
5.2	Algorithme de Baum-Welsh et problème de ré-estimation	40
5.2.1	Algorithme "forward" de Baum	40
5.2.2	Algorithme "backward" de Baum	41
5.2.3	L'algorithme de Baum-Welsh	43
5.3	Algorithme de Viterbi et problème de reconnaissance	45
6	Détection des régimes : régime crise / régime hors crise	49
6.1	Classification de données boursières	50
6.2	Problème de ré-estimation	51
6.3	Problème de reconnaissance	52
6.4	Etude des crises	53
6.5	Calcul de la VaR	54
7	Conclusion	58
	Références bibliographiques	60

Chapitre 1

Introduction

Les événements extrêmes jouent un rôle important dans l'économie mondiale. En finance, on peut citer les crashes boursiers de 1929 et de 1987. Ces événements ont une double nature : ce sont tout d'abord des événements rares, c'est à dire dont la probabilité d'occurrence est très faible ; ce sont ensuite des événements qui sont particulièrement coûteux ou dramatiques pour ceux qui les subissent. La modélisation classique ne prend pas en compte ces catastrophes et sous-estime leurs effets. Néanmoins, dans les dernières décennies, on ressent une orientation vers l'étude des événements extrêmes et la considération des queues de distribution épaisses.

La gestion de risque est la théorie qui cherche à mesurer les risques financiers et à développer des outils capables de prévoir et de limiter les grosses pertes. Les études récentes s'intéressent plutôt à la queue de la distribution des pertes d'un portefeuille. Elles convergent vers la modélisation de cette queue par des lois, dites puissances, d'où une meilleure description des événements extrêmes.

Une collaboration portant sur la gestion de risque a commencé en 2012 entre l'équipe TOSCA d'INRIA et l'entreprise ALPHABILITY. Durant cette étude, différentes modélisations pour des actifs de la bourse de Paris ont été considérées et le choix pour le modèle des lois puissances a été fait. Ce rapport de stage représente la continuation de l'étude mentionnée. L'analyse de sensibilité des paramètres du modèle par rapport à certaines perturbations représente un premier apport de ce stage. Ensuite, l'apport majeur est l'étude du changement de régime et la proposition d'une modélisation tenant compte de l'aspect crise par la définition d'une nouvelle approche de calcul de l'indicateur de risque.

Ce rapport s'organise de la façon suivante. Le chapitre 2 décrit le contexte général du stage, à savoir les différentes parties de la collaboration, l'objectif et la nature des données utilisées. Dans le chapitre 3, on rappelle la partie loi puissance tirée de l'étude précédente [8]. On détaille l'estimation des paramètres de la loi, ainsi que l'indicateur de risque et les tests de validité. Afin de mieux comprendre notre modélisation, on effectue dans le chapitre 4 une analyse de sensibilité par rapport aux perturbations extérieures.

Le faible nombre de données fourni par ALPHABALITY a limité les conclusions partant sur l'analyse de sensibilité. Cela nous a amené à raffiner notre modèle et à étudier les changements de régime afin de caractériser les données par un nouvel critère crise/hors crise. Ainsi, le chapitre 5 s'intéresse au modèle de Markov caché qu'on a utilisé pour détecter les différents régimes. Le chapitre 6 montre l'application de cette classification sur les données boursières, les résultats trouvés et la nouvelle méthode de calcul de l'indicateur Value-at-Risk en tenant compte du critère crise/non-crise. Nous terminons par un chapitre de conclusion présentant les questions ouvertes, les limitations et les perspectives de notre modèle.

Chapitre 2

Contexte du stage

La gestion de risque est au coeur de la gestion d'actifs. Le dernier krash de l'automne 2008 et la récession, qui a envahi l'économie mondiale depuis cinq ans, montrent que les techniques usuelles de mesure et de gestion de risque ne sont pas adaptées pour décrire la réalité financière, notamment car elles sous-estiment gravement les risques extrêmes. La société ALPHABILITY vise à apporter aux gérants de portefeuilles des solutions informatiques pour l'ingénierie financière. Elle propose spécifiquement des solutions à destination des sociétés de gestion de portefeuilles de petites et moyennes tailles pour leur permettre une relation personnalisée et une gestion des risques mieux sécurisée. La société ALPHABILITY inscrit la gestion de risque au coeur de son innovation de façon à proposer à ses clients des solutions efficaces à travers des indicateurs simples d'interprétation.

L'équipe TOSCA d'INRIA est spécialiste des méthodes numériques probabilistes et de la modélisation stochastique. Les mathématiques financières constituent un des thèmes majeurs d'applications étudiées par TOSCA.

Une collaboration entre ALPHABILITY et TOSCA a mené, en 2012, à une étude dont l'objectif est de définir et d'étudier des mesures fiables du risque des marchés financiers dans le but de proposer des stratégies de gestion de portefeuille tenant compte de critères de risque. Cette étude consiste à décrire le marché financier via trois modèles, à savoir le modèle gaussien, le modèle des lois stables et le modèle des lois puissances. La comparaison des différentes modélisations a montré que celle des lois puissances s'avère la plus efficace pour l'estimation des valeurs extrêmes.

Ce rapport s'inscrit dans la continuation de l'étude mentionnée. Le but est d'analyser la sensibilité du modèle choisi aux perturbations externes et aux changements de certaines valeurs d'entrée dans un premier temps, ainsi que de proposer une nouvelle méthode de calcul de la Value-at-Risk, notre indicateur de risque, dans un second temps. Cette méthode repose sur le principe de l'existence de deux régimes crise et hors crise dans le marché financier et cherche à classer les données selon ce critère et tenir compte de cette classification de nos estimateurs.

Les études numériques présentées dans ce rapport ont été réalisées à partir de données de marché fournies par ALPHABILITY. Il s'agit des cours journaliers des actions de la

bourse de Paris (Euronext Paris), avec notamment les cours à l'ouverture, à la clôture, les valeurs maximales et minimales de chaque jour de cotation. Dans ce travail, on a uniquement utilisé les cours de clôture journaliers. Les données couvrent 10 ans, de 2001 à 2011. Chaque année compte environ 252 jours d'ouverture de la bourse, ce qui conduit à un total d'environ 2520 données pour chaque actif. Notons que ces données sont loin d'être uniformes d'une action à une autre, puisque certaines données sont manquantes soit parce que les actions n'ont pas été cotées certaines jours, soit parce que les données sont incomplètes. Les actions étudiées dans ce document ont été choisies pour leur grand nombre de données, soit environ 2520.

Chapitre 3

Modélisation par les lois puissances

Dans ce chapitre, on définit le modèle des lois puissances, ainsi que les méthodes d'estimation de ses paramètres. Ce modèle s'intéresse à la queue de la distribution et donne par conséquent plus de poids aux valeurs extrêmes. À partir de cette modélisation, on cherche à construire un indicateur de mesure de risque permettant de prédire la perte maximale à subir sur un horizon temporel donné et à un niveau de confiance donné.

3.1 Introduction au modèle

Nous commençons par introduire les notions de rendements journaliers et de log-pertes.

Définition 3.1. *On considère un actif financiers de cours $(S_t, t \in \mathbb{N})$, où le temps t est mesuré par unité de jours. On appelle rendements journaliers ou plus simplement rendement de l'actif S_t , la suite $(R_t, t \in \mathbb{N})$, où*

$$R_t = \frac{S_{t+1} - S_t}{S_t}.$$

On appelle rendements logarithmiques journaliers ou log-rendements la suite $(L_t, t \in \mathbb{N})$ où

$$L_t = \ln(S_{t+1}) - \ln(S_t) = \ln\left(\frac{S_{t+1}}{S_t}\right).$$

Enfin, les log-pertes correspondent aux valeurs $-L_t$ lorsque $L_t < 0$.

Dans tout ce rapport, nous allons utiliser ces notions de log-rendements et de log-pertes.

Selon l'étude précédente, la modélisation par la loi normale et les lois α -stables [26] permet de décrire le comportement de toute la distribution, mais manque de précision sur l'évaluation des événements rares. En effet, la loi normale sous-estime la VaR alors que les lois α -stables la sur-estiment. Dans cette perspective, on s'intéresse à un modèle qui se focalise sur la queue de la distribution. Il s'agit des lois puissance de la forme

$$1 - F(x) = \ell(x)x^{-\alpha} \tag{3.1}$$

où

- la fonction $F(x) = \mathbb{P}(X \leq x)$ est la fonction de répartition de la distribution des log-pertes des actifs financiers ;
- la fonction $\ell(x)$ est une fonction à variation lente ;
- le paramètre $\alpha > 0$ est l'indice caractéristique de la décroissance de la fonction $1 - F(\cdot)$.

Définition 3.2. Une fonction mesurable ℓ de \mathbb{R}_+ dans \mathbb{R}_+ est dite à variation lente, au voisinage de l'infini, si pour tout $x > 0$,

$$\lim_{t \rightarrow +\infty} \frac{\ell(tx)}{\ell(t)} = 1 .$$

On note \mathcal{R}_0 l'ensemble des fonctions à variation lente.

Dans les parties qui suivent, on s'intéresse à l'estimation de l'indice de queue α par l'estimateur de Hill. Dans un premier temps, on étudie son comportement asymptotique. Ensuite, on propose une approximation pour la fonction à variations lentes $\ell(x)$. Enfin, on définit l'indicateur de risque VaR et on l'évalue à l'aide de cet estimateur.

3.2 Estimateur de Hill

Dans cette partie, on cherche à donner une estimation de l'indice de queue de la distribution des rendements financiers sous le modèle de loi puissance

$$1 - F(x) = \ell(x)x^{-\alpha}.$$

On s'intéressera principalement dans la suite au cas où $\ell(x)$ converge vers une constante positive C quand x tend vers l'infini.

Définition 3.3. Soit X une variable aléatoire réelle. La fonction quantile et la fonction queue sont définies respectivement par :

$$Q(y) = F^{-1}(y), \quad y \in [0, 1],$$

et

$$U(x) = Q\left(1 - \frac{1}{x}\right), \quad x \in [1, +\infty),$$

où $F^{-1}(x) = \inf\{u, F(u) \geq x\}$, noté la fonction pseudo-inverse de F .

Pour tout $0 < y < 1$, on a alors

$$y = Q(1 - y)^{-\alpha} \ell(Q(1 - y)). \quad (3.2)$$

Si on pose $\gamma = \frac{1}{\alpha}$, la fonction queue vérifie

$$U(x) = x^\gamma \ell^*(x), \quad x > 1 \quad (3.3)$$

et le passage au log nous donne

$$\log(Q(1-y)) = -\gamma \log(y) + \log \ell^*\left(\frac{1}{y}\right), \quad 0 < y < 1, \quad (3.4)$$

où ℓ^* est encore une fonction à variation lente définie sur $]1, \infty[$ et reliée à ℓ via la conjugaison de Bruyn [6], p.79. On peut par exemple vérifier que ℓ^* est à variation lente en utilisant la représentation de Karamata [24].

On considère les données suivantes :

- La série de log-pertes journalières observées sur le marché est $(X_i)_{1 \leq i \leq n}$. Ici, les X_i sont positifs, et les plus grandes valeurs correspondent aux pertes les plus importantes.
- Les statistiques d'ordre de cette série est $(X_{(i)})_{1 \leq i \leq n}$ telle que $X_{(i)} \leq X_{(i+1)}$ pour tout $1 \leq i \leq n-1$.

Une propriété classique assure que, pour toute suite i_n telle que $i_n/n \rightarrow y$, $X_{(i_n)}$ est un estimateur consistant pour $Q(y)$ lorsque $n \rightarrow \infty$. Autrement dit, pour i et n fixés, on peut voir $X_{(i)}$ comme une approximation de $Q(\frac{i}{n+1})$.

Cette observation permet de vérifier graphiquement à partir de (3.4) la linéarité de la fonction log-quantile. On trace le graphe de Pareto (Pareto Qantile Plot), construit comme suit : Pour des valeurs de la série $(X_{(i)})_{1 \leq i \leq n}$, on trace $\log X_{(n+1-i)}$ en ordonnée contre $-\log(1 - i/(n+1))$ en abscisse. Si les données suivent une loi puissance, un tel graphe doit se comporter de façon linéaire.

Le graphe de Pareto, obtenu à partir des log-pertes journalières de l'action BNP Paribas sur la période 2001-2011, est donné par la figure 3.1. On observe sur cette figure que la fonction log-quantile a l'allure d'une droite sur la queue de distribution, c'est-à-dire pour les grandes valeurs de $-\log(1 - i/(n+1))$. Déterminer un estimateur de γ (et donc de α) revient à déterminer la pente de cette partie approximativement linéaire correspondant aux valeurs extrêmes.

Parmi les estimateurs les plus classiques de cette pente, on trouve celui de Hill introduit en 1975 [18], qui reste un des plus utilisé sur les marchés financiers vu son efficacité et sa simplicité. On montre dans ce qui suit qu'il s'agit d'un estimateur consistant qui converge en probabilité vers α .

Définition 3.4. *On définit l'estimateur de Hill par*

$$\hat{\alpha}_r = \left[r^{-1} \sum_{i=1}^r \log(X_{(n-i+1)}) - \log(X_{(n-r)}) \right]^{-1}, \quad (3.5)$$

pour un $0 < r < n$ fixé.

Afin d'alléger les notations, lorsqu'il n'y a pas d'ambiguïté sur le choix de r , on utilisera la notation $\hat{\alpha}$. La consistance de cet estimateur découle du résultat suivant.

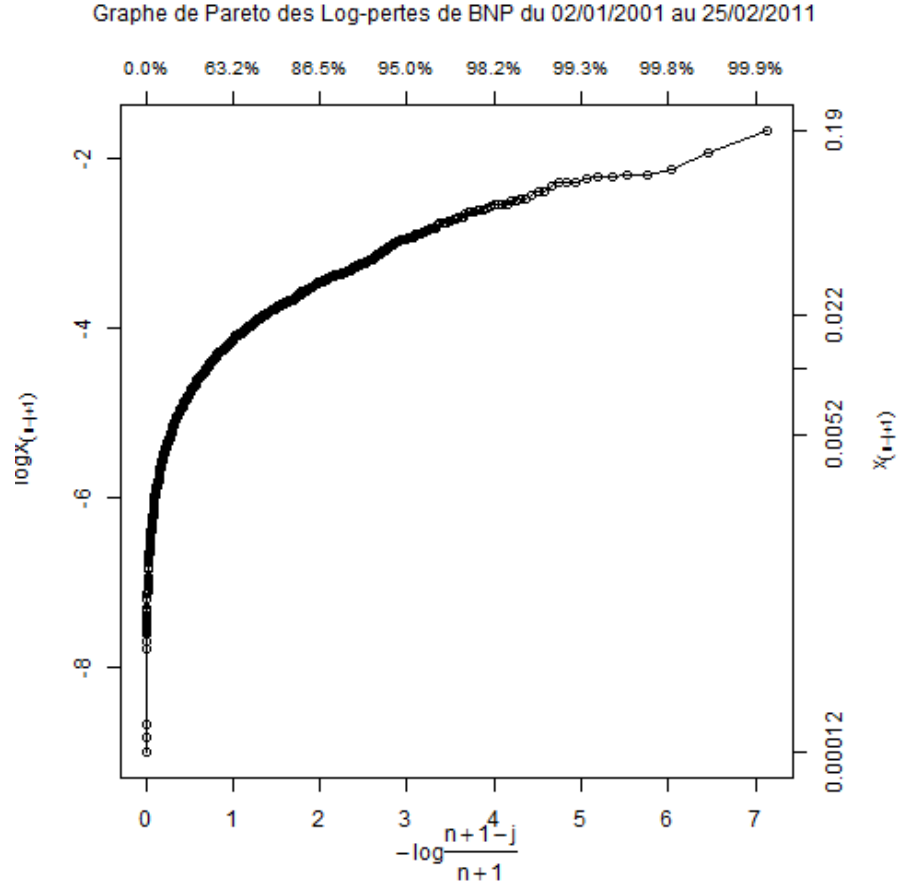


FIGURE 3.1 – Graphe de Pareto des log-pertes de l'actif BNP Paribas sur la période Jan. 2001-Fév. 2011

Théorème 3.1. *On définit le vecteur aléatoire de log-scaling $\{Z_j; 1 \leq j \leq r < n\}$ par :*

$$Z_j = j(\log X_{(n-j+1)} - \log X_{(n-j)}). \quad (3.6)$$

On a, alors :

$$(Z_1, \dots, Z_r) \stackrel{\mathcal{L}}{=} \gamma(E_1, \dots, E_r) + (\xi_1, \dots, \xi_r) \quad (3.7)$$

où $\stackrel{\mathcal{L}}{=}$ désigne l'égalité en loi, (E_1, \dots, E_r) est un vecteur de variables aléatoires i.i.d. de loi exponentielle de paramètre 1 et (ξ_1, \dots, ξ_r) est un vecteur de termes d'erreur tel que

$$\xi(r; n) = \frac{1}{r} \sum_{j=1}^r \xi_j, \quad (3.8)$$

converge en probabilité vers 0 lorsque $n, r \rightarrow +\infty$ et $\frac{r}{n} \rightarrow 0$.

Remarquons que l'estimateur de Hill peut se ré-écrire sous la forme

$$\hat{\alpha}^{-1} = \frac{1}{r} \sum_{j=1}^r Z_j. \quad (3.9)$$

Ainsi, le corollaire suivant du théorème 3.1 nous donne la consistance.

Corollaire 3.1. *Soit $(r_n)_{n \geq 1}$ une suite d'entiers positifs tels que $r_n \rightarrow +\infty$ et $\frac{r_n}{n} \rightarrow 0$ lorsque $n \rightarrow +\infty$. Nous avons la convergence en probabilité*

$$\left(r_n^{-1} \sum_{i=1}^{r_n} \log(X_{(n-i+1)}) - \log(X_{(n-r_n)}) \right) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \gamma = \frac{1}{\alpha}. \quad (3.10)$$

Preuve du corollaire 3.1. L'estimateur de Hill donné en (3.9) s'écrit, d'après le théorème 3.1,

$$\hat{\alpha}^{-1} = \gamma \cdot \frac{1}{r_n} \sum_{j=1}^{r_n} E_j + \xi(r_n; n),$$

avec $\xi(r_n; n) \rightarrow 0$ en probabilité lorsque $n \rightarrow +\infty$. Or, par la loi des grands nombres,

$$\frac{1}{r_n} \sum_{j=1}^{r_n} E_j \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}(E_1) = 1$$

presque sûrement, donc elle converge en probabilité. Ainsi, $\hat{\alpha}^{-1}$ converge en probabilité vers γ lorsque $n \rightarrow +\infty$. \square

Dans la suite, on donnera quelques résultats utiles pour démontrer le théorème 3.1. Ces résultats ont été étudiés dans mon projet d'initiation à la recherche du master 2 MFA [25]. On introduira l'approche de Rényi qui présente une manière d'écrire l'estimateur de Hill par l'utilisation de la représentation de Rényi (1953) [23] des statistiques d'ordre des v.a. de loi exponentielle.

Proposition 3.1. *Soit $\{E_i; 1 \leq i \leq n\}$ un vecteur de variables aléatoires i.i.d. de loi exponentielle de paramètre λ , et soient $\{E_{(i)}; 1 \leq i \leq n\}$ les statistiques d'ordre correspondantes.*

Alors $(E_{(1)}, E_{(2)} - E_{(1)}, \dots, E_{(n)} - E_{(n-1)})$ sont indépendantes et suivent des lois exponentielles de paramètres respectifs $n\lambda, (n-1)\lambda, \dots, \lambda$.

Preuve. Soit f une fonction bornée. Nous pouvons alors écrire

$$\begin{aligned} & \mathbb{E}(f(E_{(1)}, E_{(2)} - E_{(1)}, \dots, E_{(n)} - E_{(n-1)})) \\ &= n! \mathbb{E}[f(E_1, E_2 - E_1, \dots, E_n - E_{n-1}); E_1 < E_2 < \dots < E_n] \\ &= n! \int \dots \int_{0 < x_1 < x_2 < \dots < x_n} f(x_1, x_2 - x_1, \dots, x_n - x_{n-1}) e^{-\lambda \sum_{i=1}^n x_i} \lambda^n dx_1 \dots dx_n. \end{aligned}$$

En effectuant le changement de variables

$$y_i = x_i - x_{i-1} \text{ pour } 2 \leq i \leq n \text{ et } y_1 = x_1,$$

et en remarquant que $\|\det J\|$, le déterminant de la Jacobienne de ce changement de variables, vaut 1, on déduit que

$$\begin{aligned}
& \mathbb{E}(f(E_{(1)}, E_{(2)} - E_{(1)}, \dots, E_{(n)} - E_{(n-1)})) \\
&= n! \int_0^\infty \int_0^\infty \dots \int_0^\infty f(y_1, y_2, \dots, y_n) e^{-\lambda \sum_{j=1}^n j y_{n-j+1}} \lambda^n dy_1 \dots dy_n \\
&= n! \int_0^\infty \int_0^\infty \dots \int_0^\infty f(y_1, y_2, \dots, y_n) e^{-\lambda \sum_{i=1}^n ((n-i+1)y_i)} \lambda^n dy_1 \dots dy_n \\
&= \int_0^\infty \int_0^\infty \dots \int_0^\infty f(y_1, y_2, \dots, y_n) \prod_{i=1}^n (n-i+1) \lambda e^{-(n-i+1)\lambda y_i} dy_1 \dots dy_n.
\end{aligned}$$

On conclut que $(E_{(1)}, E_{(2)} - E_{(1)}, \dots, E_{(n)} - E_{(n-1)})$ sont indépendantes et suivent des lois exponentielles de paramètres respectivement $n\lambda, (n-1)\lambda, \dots, \lambda$. □

Théorème 3.2 (théorème de représentation de A. Rényi [23]). *Soit $\{E_i; 1 \leq i \leq n\}$ un vecteur de variables aléatoires i.i.d. de loi exponentielle de paramètre 1 et soit $\{E_{(i)}; 1 \leq i \leq n\}$ les statistiques d'ordre correspondantes, alors*

$$E_{(n-j+1)} - E_{(n-r)} \stackrel{\mathcal{L}}{=} \sum_{i=j}^r \frac{E'_i}{i}, \quad \text{pour tout } 1 \leq j \leq r < n, \quad (3.11)$$

où $\{E'_i; 1 \leq i \leq n\}$ est encore un vecteur de variables aléatoires i.i.d. de loi exponentielle de paramètre 1 et $\stackrel{\mathcal{L}}{=}$ désigne l'égalité en loi.

Preuve. On commence par remarquer que, pour tout $1 \leq j \leq r < n$,

$$E_{(n-j+1)} - E_{(n-r)} = \sum_{i=j}^r (E_{n-i+1,n} - E_{n-i,n}).$$

D'autre part, d'après la proposition 3.1, $E_{(1)}, E_{(2)} - E_{(1)}, \dots, E_{(n)} - E_{(n-1)}$ sont indépendantes et suivent des lois exponentielles de paramètres respectivement $n, (n-1), \dots, 1$. D'où,

$$(E_{(1)}, E_{(2)} - E_{(1)}, \dots, E_{(n)} - E_{(n-1)}) \stackrel{\mathcal{L}}{=} \left(\frac{E'_1}{n}, \frac{E'_2}{n-1}, \dots, \frac{E'_n}{1} \right),$$

où $\{E'_i; 1 \leq i \leq n\}$ est encore un vecteur de variables aléatoires i.i.d. de loi exponentielle de paramètre 1.

Ainsi, pour tout $1 \leq j \leq r < n$,

$$E_{(n-j+1)} - E_{(n-r)} = \sum_{i=j}^r (E_{n-i+1,n} - E_{n-i,n}) \stackrel{\mathcal{L}}{=} \sum_{i=j}^r \frac{E'_i}{i}.$$

Ce qui conclut la démonstration. \square

Corollaire 3.2. *Pour tout $1 \leq j \leq r+1$,*

$$\begin{cases} \mathbb{E}(E_{(n-j+1)}) = \log(\frac{n}{j}) + o(1) \rightarrow +\infty, \\ \mathbb{P}\left(|E_{(n-j+1)} - \mathbb{E}(E_{(n-j+1)})| > \sqrt{\log(\frac{n}{j})}\right) \rightarrow 0, \end{cases} \quad (3.12)$$

lorsque $n, r \rightarrow +\infty$ et $\frac{r}{n} \rightarrow 0$. D'où, $E_{(n-j+1)} \rightarrow +\infty$ en probabilité lorsque $n, r \rightarrow +\infty$ et $\frac{r}{n} \rightarrow 0$.

Preuve. En utilisant le théorème de représentation de Rényi, on obtient :

$$E_{(n-j+1)} - E_{(1)} \stackrel{\mathcal{L}}{=} \sum_{i=j}^{n-1} \frac{E'_i}{i}, \quad \text{pour tout } 1 \leq j < n.$$

Or d'après la proposition 3.1, on a $\mathbb{E}(E_{(1)}) = \frac{1}{n}$ et $\text{Var}(E_{(1)}) = \frac{1}{n^2}$. Ainsi, par passage à l'espérance et à la variance, on obtient, pour tout $1 \leq j < n$

$$\mathbb{E}(E_{(n-j+1)}) = \sum_{i=j}^n \frac{1}{i} = \log(\frac{n}{j}) + o(1),$$

et

$$\text{Var}(E_{(n-j+1)}) = \sum_{i=j}^n \frac{1}{i^2} \leq c = \frac{\pi^2}{6}.$$

On conclut par l'inégalité de Tchebychev que

$$\mathbb{P}\left(|E_{(n-j+1)} - \mathbb{E}(E_{(n-j+1)})| > \sqrt{\log(\frac{n}{j})}\right) \leq \frac{\text{Var}(E_{(n-j+1)})}{\log(\frac{n}{j})} \leq \frac{c}{\log(\frac{n}{j})}.$$

Nous avons donc prouvé (3.12).

De plus,

$$\begin{aligned} \mathbb{P}(E_{(n-j+1)} < \log(\frac{n}{j}) - \sqrt{\log(\frac{n}{j})}) &\leq \mathbb{P}\left(\left\{E_{(n-j+1)} < \log(\frac{n}{j}) - \sqrt{\log(\frac{n}{j})}\right\} \right. \\ &\quad \left. \cup \left\{E_{(n-j+1)} > \log(\frac{n}{j}) + \sqrt{\log(\frac{n}{j})}\right\}\right) \\ &\leq \mathbb{P}\left(|E_{(n-j+1)} - \mathbb{E}(E_{(n-j+1)})| > \sqrt{\log(\frac{n}{j})}\right) \rightarrow 0, \end{aligned}$$

lorsque $n, r \rightarrow +\infty$ et $\frac{r}{n} \rightarrow 0$. Ainsi, $E_{(n-j+1)} \rightarrow +\infty$ en probabilité lorsque $n, r \rightarrow +\infty$ et $\frac{r}{n} \rightarrow 0$. \square

Preuve du théorème 3.1. Soit $\{E_i; 1 \leq i \leq n\}$ un vecteur de variables aléatoires i.i.d. de loi exponentielle de paramètre 1. On note $F_E(x) = 1 - e^{-x}$, $x > 0$ sa fonction de répartition. Soient $\{E_{(i)}; 1 \leq i \leq n\}$ les statistiques d'ordre correspondantes. On commence dans un premier temps par montrer que

$$\{X_{(j)}; 1 \leq j \leq n\} \stackrel{\mathcal{L}}{=} \{U(e^{E_{(j)}}); 1 \leq j \leq n\}.$$

Par double utilisation du théorème des fonctions de répartition inverses, on obtient dans un premier temps, que pour tout $1 \leq j \leq n$, $F_E(E_j) = 1 - e^{-E_j}$ est de loi uniforme sur $[0,1]$. Puis, dans un deuxième temps, on déduit que, pour tout $1 \leq j \leq n$,

$$X_j \stackrel{\mathcal{L}}{=} F^{-1}(1 - e^{-E_j}) \stackrel{\mathcal{L}}{=} U(E_j).$$

Comme la fonction exponentielle et la fonction quantile de queue U sont croissantes, on obtient, par passage aux statistiques d'ordre,

$$\{X_{(j)}; 1 \leq j \leq n\} \stackrel{\mathcal{L}}{=} \{U(e^{E_{(j)}}); 1 \leq j \leq n\}. \quad (3.13)$$

Revenons à la démonstration de l'égalité (3.7). On a

$$\begin{aligned} \{Z_j; 1 \leq j \leq r\} &= \{j(\log \frac{X_{(n-j+1)}}{X_{(n-j)}}); 1 \leq j \leq r\} \\ &\stackrel{\mathcal{L}}{=} \{j(\log \frac{U(e^{E_{(n-j+1)}})}{U(e^{E_{(n-j)}})}); 1 \leq j \leq r\}. \end{aligned}$$

D'après le théorème 3.2 de représentation de Rényi, $E_{(n-j+1)} - E_{(n-j)} = \frac{E'_j}{j}$ où E'_j est une variable aléatoire de loi exponentielle de paramètre 1, indépendante de $E_{(n-j)}$ par la proposition 3.2. Ainsi

$$\begin{aligned} \{Z_j; 1 \leq j \leq r\} &\stackrel{\mathcal{L}}{=} \{j(\log \frac{U(e^{\frac{E'_j}{j}} e^{E_{(n-j)}})}{U(e^{E_{(n-j)}})}); 1 \leq j \leq r\} \\ &\stackrel{\mathcal{L}}{=} \{j(\log[(e^{\frac{E'_j}{j}})^{\gamma} \frac{\ell_U(e^{\frac{E'_j}{j}} e^{E_{(n-j)}})}{\ell_U(e^{E_{(n-j)}})}]); 1 \leq j \leq r\}, \end{aligned}$$

d'après l'équation (3.3) définissant notre fonction quantile de queue de loi de type Pareto. On conclut que

$$\{Z_j; 1 \leq j \leq r\} \stackrel{\mathcal{L}}{=} \{\gamma E'_j + j \left(\log \frac{\ell_U(e^{E_{(n-j+1)}})}{\ell_U(e^{E_{(n-j)}})} \right); 1 \leq j \leq r\}.$$

On note, pour tout $1 \leq j \leq r$,

$$\xi_{j,n} = j \left(\log \frac{\ell_U(e^{E_{(n-j+1)}})}{\ell_U(e^{E_{(n-j)}})} \right). \quad (3.14)$$

Montrons ensuite que $\xi(r; n) = \frac{1}{r} \sum_{j=1}^r \xi_{j,n}$ converge en probabilité vers 0 lorsque $n, r \rightarrow +\infty$ et $\frac{r}{n} \rightarrow 0$. On a

$$\begin{aligned}\xi(r; n) &= \frac{1}{r} \sum_{j=1}^r j \log \frac{\ell_U(e^{E_{(n-j+1)}})}{\ell_U(e^{E_{(n-j)}})} \\ &= \frac{1}{r} \sum_{j=1}^r \log \frac{\ell_U(e^{E_{(n-j+1)}})}{\ell_U(e^{E_{(n-r)}})}.\end{aligned}$$

Par suite, la représentation de Karamata [24] des fonctions à variation lente nous donne

$$\xi(r; n) = \frac{1}{r} \sum_{j=1}^r \log \frac{c(e^{E_{(n-j+1)}})}{c(e^{E_{(n-r)}})} + \frac{1}{r} \sum_{j=1}^r \int_{e^{E_{(n-r)}}}^{e^{E_{(n-j+1)}}} \frac{\varepsilon(u)}{u} du$$

avec $c(x) \rightarrow c \in (0, +\infty)$ et $\varepsilon(x) \rightarrow 0$ lorsque $x \rightarrow +\infty$.

Ainsi, on obtient

$$|\xi(r; n)| \leq \left| \frac{1}{r} \sum_{j=1}^r \log \frac{c(e^{E_{(n-j+1)}})}{c(e^{E_{(n-r)}})} \right| + \sup_{u \geq e^{E_{(n-r)}}} |\varepsilon(u)| \cdot \left| \frac{1}{r} \sum_{j=1}^r (E_{(n-j+1)} - E_{(n-r)}) \right|.$$

Comme

$$\frac{1}{r} \sum_{j=1}^r (E_{(n-j+1)} - E_{(n-r)}) = \frac{1}{r} \sum_{j=1}^r \sum_{i=j}^r \frac{E'_i}{i} = \frac{1}{r} \sum_{j=1}^r E'_j,$$

on déduit que

$$|\xi(r; n)| \leq \left| \frac{1}{r} \sum_{j=1}^r \log \frac{c(e^{E_{(n-j+1)}})}{c(e^{E_{(n-r)}})} \right| + \sup_{u \geq e^{E_{(n-r)}}} |\varepsilon(u)| \cdot \left| \frac{1}{r} \sum_{j=1}^r E'_j \right|.$$

Ensuite, en passant à la limite lorsque $n, k \rightarrow \infty$ et $\frac{k}{n} \rightarrow 0$, on obtient

- par la loi des grands nombres, $\frac{1}{r} \sum_{j=1}^r E'_j \rightarrow \mathbb{E}(E'_1) = 1$ presque sûrement, donc en probabilité,
- d'après le corollaire 3.2, $e^{E_{(n-r)}} \rightarrow +\infty$ en probabilité, donc $\sup_{u \geq e^{E_{(n-r)}}} |\varepsilon(u)| \rightarrow 0$ en probabilité,
- d'après le corollaire 3.2, $e^{E_{(n-r)}}$ et $e^{E_{(n-j+1)}}$ tendent vers $+\infty$ en probabilité, donc $c(e^{E_{(n-j+1)}})$ et $c(e^{E_{(n-r)}})$ convergent vers c en probabilité, donc par l'application du théorème de Cesàro, $\frac{1}{k} \sum_{j=1}^k \log \frac{c(e^{E_{(n-j+1)}})}{c(e^{E_{(n-r)}})}$ converge en probabilité, et sa limite est 0.

D'où le résultat, $\xi(r; n)$ converge en probabilité vers 0. \square

Remarque 3.1. On constate qu'une grande latitude est laissée pour le choix du paramètre r_n dans le résultat précédent. En pratique, le choix de ce paramètre doit être un compromis entre une petite valeur qui donnerait un nombre insuffisant de données pour estimer α , et une valeur trop grande qui incluerait dans l'estimateur des valeurs qui ne sont pas dans la queue de distribution.

Afin de construire des intervalles de confiance sur l'estimation de α , par l'estimateur de Hill, le théorème de la limite centrale suivant est utilisé.

Théorème 3.3. *On considère $r_n \in \mathbb{N}$ tel que $r_n \rightarrow \infty$ et $r_n = o(n)$ lorsque $n \rightarrow \infty$. On pose :*

$$I(n, r_n, x) = \left(\frac{r_n}{n}\right) \left(1 + \frac{x}{r_n^{1/2}}\right),$$

et

$$J(F, n, r_n, x) = r_n^{1/2} \int_1^\infty \left(y^{-\alpha} - \frac{1 - F(F^{-1}(1 - I(n, r_n, x))y)}{I(n, r_n, x)} \right) \frac{dy}{y}.$$

Si $J(F, n, r_n, x) \xrightarrow[n \rightarrow \infty]{} A \in \mathbb{R}$ uniformément sur tout ensemble de compact, alors

$$r_n^{1/2}(\hat{\alpha}^{-1} - \alpha^{-1}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(-A, \alpha^{-2}),$$

où $\hat{\alpha}$ est l'estimateur de Hill du paramètre α .

La preuve de ce théorème se trouve dans [16].

Remarque 3.2. *Ce théorème permet d'obtenir l'intervalle de confiance de l'estimateur de Hill à un niveau de confiance de $(1 - \beta)$:*

$$\alpha^{-1} \in \left[\hat{\alpha}^{-1} + \left(-A - \frac{\Phi(1 - \beta/2)}{\hat{\alpha}} \right) \frac{1}{r_n^{1/2}}, \hat{\alpha}^{-1} + \left(-A + \frac{\Phi(1 - \beta/2)}{\hat{\alpha}} \right) \frac{1}{r_n^{1/2}} \right],$$

où Φ est la fonction quantile de la loi normale centrée réduite.

Remarque 3.3. *Lorsque la fonction à variation lente $\ell(x)$ est égale à une constante $C > 0$ pour x suffisamment grand (et également lorsque $\ell(x)$ converge vers $C > 0$ suffisamment vite), la constante A dans le théorème 3.3 est nulle [6]. Nous supposons implicitement dans la suite que les actifs du marché financier satisfont ces hypothèses.*

3.3 Estimation de la fonction à variations lentes et des quantiles

Nous nous placerons dans cette section sous l'hypothèse que la fonction à variation lente $\ell(x)$ converge vers une constante $C > 0$ lorsque $x \rightarrow +\infty$.

Rappelons la relation (3.2) qui relie la fonction quantile à la fonction à variation lente : $1 - p = Q(p)^{-\alpha} \ell(Q(p))$. Une propriété classique assure que, pour toute suite i_n telle que $i_n/n \rightarrow p$, $X_{(i_n)}$ est un estimateur consistant pour $Q(p)$ lorsque $n \rightarrow \infty$. Autrement dit, pour i et n fixés, on peut voir $X_{(i)}$ comme une approximation de $Q(\frac{i}{n+1})$.

Ainsi, pour p proche de 1 tel que $\ell(Q(p))$ est correctement approchée par sa limite C , choisissons un entier $w_n \leq n$ tel que $(w_n + 1)/(n + 1) \approx 1 - p$. Par la suite, en utilisant

l'approximation de $Q(p)$ par $X_{(n-w_n)}$ et celle de α par son estimateur de Hill, nous obtenons que la constante C est approchée par

$$\hat{C} \approx X_{(n-w_n)}^{\hat{\alpha}_{r_n}} \frac{w_n + 1}{n + 1}. \quad (3.15)$$

On déduit, par conséquent, l'estimateur suivant pour le quantile de niveau p :

$$\hat{q}_{w_n, r_n, p} = X_{(n-w_n)} \left(\frac{w_n + 1}{(n + 1)p} \right)^{\hat{\alpha}_{r_n}^{-1}}, \quad (3.16)$$

où $\hat{\alpha}_{r_n}$ désigne l'estimateur de Hill obtenu avec le choix $r = r_n$. Ces estimateurs ont été proposés initialement par Weissman [27].

Afin de vérifier la consistance de cet estimateur d'un quantile et de construire un intervalle de confiance, nous allons utiliser l'écriture (3.3) de la fonction queue. Remarquons que $Q(1 - p) = U(1/p) = p^{-\gamma} \ell^*(1/p)$. Ainsi, par utilisation du théorème des fonctions de répartition inverses : $X_{(n-w_n)}$ a la même loi que $Q(1 - U_{(w_n+1)}) = U_{(w_n+1)}^{-\gamma} \ell^*(U_{(w_n+1)}^{-1})$, où $U_{(1)}, \dots, U_{(n)}$ sont les statistiques d'ordre d'une famille i.i.d de n v.a. uniformes sur $[0, 1]$. En utilisant ces deux relations, nous obtenons

$$\log \frac{\hat{q}_{w_n, r_n, p}}{Q(1 - p)} = \gamma \left(-\log U_{(w_n+1)} - \log \frac{n + 1}{w_n + 1} \right) + (\hat{\alpha}_{r_n}^{-1} - \gamma) \log \left(\frac{w_n + 1}{(n + 1)p} \right) + \log \frac{\ell^*(U_{(w_n+1)}^{-1})}{\ell^*(1/p)}.$$

Il est connu [6] que

$$\sqrt{w_n} \left(-\log U_{(w_n+1)} - \log \frac{n + 1}{w_n + 1} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

lorsque $n \rightarrow +\infty$, $w_n \rightarrow +\infty$ et $w_n/n \rightarrow 0$. Le théorème 3.3 donne la loi limite de $\sqrt{r_n}(\hat{\alpha}_{r_n}^{-1} - \gamma)$. Le seul terme dont on ne peut pas estimer le comportement asymptotique en toute généralité est le dernier. Cependant, il est nul si $\ell^*(x) = C > 0$ pour x assez grand (ou si $\ell^*(x)$ converge vers $C > 0$ suffisamment rapidement). De nouveau, nous nous placerons dans ce cadre.

Sous cette hypothèse, il est alors possible de construire un intervalle de confiance pour $Q(1 - p)$. Par exemple, si on choisit $w_n = r_n$, en se souvenant que le biais asymptotique de l'estimateur de Hill (la constante A du théorème 3.3) est nul sous notre dernière hypothèse, nous obtenons

$$\sqrt{r_n} \left(1 + \log^2 \frac{r_n + 1}{(n + 1)p} \right)^{-1/2} \left(\frac{\hat{q}_{r_n, r_n, p}}{Q(1 - p)} - 1 \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^2),$$

quand $n \rightarrow +\infty$, $r_n \rightarrow +\infty$ et $r_n/n \rightarrow 0$. On en déduit l'intervalle de confiance pour $Q(1 - p)$ de niveau asymptotique $1 - \beta$

$$\left[\frac{\hat{q}_{r_n, r_n, p}}{1 + \Phi^{-1}(1 - \beta/2) \frac{\hat{\alpha}_{r_n}^{-1}}{\sqrt{k}} \sqrt{1 + \log^2 \frac{r_n + 1}{(n + 1)p}}}, \frac{\hat{q}_{r_n, r_n, p}}{1 - \Phi^{-1}(1 - \beta/2) \frac{\hat{\alpha}_{r_n}^{-1}}{\sqrt{k}} \sqrt{1 + \log^2 \frac{r_n + 1}{(n + 1)p}}} \right].$$

3.4 La *Value-at-Risk* et le modèle des lois puissances

La *Value-at-Risk* est un indicateur popularisé par la banque JP Morgan en 1993 et utilisé pour mesurer le risque du marché d'un portefeuille d'instruments financiers. Cet indicateur est défini comme la perte maximale potentielle qui ne devrait être atteinte qu'avec une probabilité donnée p sur un horizon temporel donné. Le principal attrait de cette mesure réside dans le fait de permettre de quantifier le risque et de proposer une mesure homogène, générale et compréhensible par tous.

Définition 3.5. La *Value-at-Risk*, notée VaR , d'un titre financier de cours $(S_t, t \geq 0)$ sur un horizon temporel T (exprimé en jours) et à un seuil de confiance a est la valeur absolue du $(1 - a)^{\text{ème}}$ quantile de la distribution de profits et de pertes sur une durée T , c'est-à-dire

$$1 - a = \mathbb{P}[S_T - S_0 < -VaR_a(T)]. \quad (3.17)$$

Lorsque l'horizon de temps T est 1 jour, cette définition s'exprime à l'aide du rendement journalier :

$$1 - a = \mathbb{P} \left[R_1 < -\frac{VaR_a(1)}{S_0} \right],$$

et à l'aide du log-rendement journalier comme suit :

$$1 - a = \mathbb{P} \left[L_1 < \log \left(1 - \frac{VaR_a(1)}{S_0} \right) \right]. \quad (3.18)$$

La VaR s'exprime en unité monétaire.

En pratique, on cherche donc la plus grande valeur ℓ (généralement négative) telle que

$$\mathbb{P}[L_1 < \ell] \leq 1 - a,$$

ce qui donne pour ℓ le $(1 - a)^{\text{ème}}$ quantile de L_1 , que l'on va estimer à partir des données. On utilise alors la formule $VaR_a(1) = S_0(1 - \exp(\ell))$ pour calculer la VaR .

En pratique, on cherche le montant tel que, 99 fois sur 100, les pertes lui soient inférieures. On parle de VaR à un niveau de confiance de 99%.

Proposition 3.2. Dans le modèle des lois puissances, la VaR au seuil de confiance a à un horizon t , est donnée par la relation

$$\log \left(1 - \frac{VaR_a(t)}{S_0} \right) = q_{1-a},$$

où q_{1-a} est le $(1 - a)^{\text{ème}}$ quantile de la loi puissance qui modélise les log-pertes journalières.

Soit $(X_i)_{1 \leq i \leq n}$ une série de log-pertes journalières observées sur le marché. On peut estimer la VaR d'horizon 1 jour à partir de la formule précédente en utilisant les estimateurs de Weissman (3.15) et (3.16) :

$$\log \left(1 - \frac{\hat{VaR}_a(t)}{S_0} \right) = \hat{q}_{1-a} = \left(\frac{\hat{C}}{1 - a} \right)^{\hat{\alpha}^{-1}}, \quad (3.19)$$

où $\hat{\alpha}$ désigne l'estimateur de Hill de l'indice de queue α . Ainsi, un estimateur de la VaR est donné par :

$$\hat{VaR}_a(t) = S_0 \left(1 - \exp \left(\frac{\hat{C}}{1-a} \right)^{\hat{\alpha}-1} \right). \quad (3.20)$$

3.5 Etude numérique : estimation des paramètres de la loi puissance, calcul de la VaR

L'étude numérique est basée sur les données de marché fournies par ALPHABILITY. Il est à mentionner que ces données ne couvrent que dix années, d'où un faible nombre de données disponibles pour faire nos estimations. Il est inutile de chercher une méthode très fine, puisque les erreurs statistiques seront dans tous les cas importantes. Vu qu'on cherche un bon compromis entre précision et coût numérique, on va opter pour une méthode d'estimation simple inspirée de la discussion ci-dessus.

Nous considérons ici l'estimation des paramètres de la loi puissance sur le cours de l'actif BNP Parisbas, en considérant des fenêtres de 252 jours.

L'estimation de l'indice $\gamma = 1/\alpha$ est donnée par le calcul de la pente sur le graphe de Pareto (figure 3.1) avec une technique de moindres carrés sur un intervalle.

Ainsi, on considère que pour $i \in [i_0, i_1]$,

$$\log X_{(i)} = -\gamma \log \left(\frac{i}{n+1} \right) + K + \epsilon_i$$

où ϵ_i est le terme d'erreur, K une constante,

$$i_0 = \lfloor p_0 n \rfloor \text{ et } i_1 = \lfloor p_1 n \rfloor \text{ avec } p_0 = 0.95 \text{ et } p_1 = 0.99.$$

Comme déjà vu, $X_{(i_j)} \approx Q(i_j/n) \approx Q(p_j)$ pour $j = 0, 1$. Ainsi, γ et K sont estimés comme la pente $\hat{\gamma}^{LS}$ et la constante \hat{K}^{LS} qui minimisent $\sum_{i=i_0}^{i_1} \epsilon_i^2$. Nous posons, par la suite, $\hat{\alpha}^{LS} = 1/\hat{\gamma}^{LS}$. Ce choix ne correspond pas à l'estimateur de Hill. Cependant, ce dernier représente une version pondérée de l'estimateur moindres carrés, et nous utilisons simplement une version non pondérée. Comme déjà dit, on justifie ce choix par la ressemblance des résultats des deux méthodes due aux erreurs statistiques et au faible nombre de données.

Il est à noter que, sur une fenêtre de 252 jours, seule environ la moitié représente les log-pertes, et donc entre 5 et 10 données seulement sont utilisés pour estimer la queue de distribution. Ainsi, notre choix de ne pas considérer les quantiles supérieurs à 99% s'avère pertinent pour éviter que les valeurs les plus grandes perturbent trop l'estimation de γ .

L'estimation de la constante limite de la fonction à variations lentes est donnée par l'estimateur de Weissman, en prenant $\hat{C} = X_{(\lfloor nw \rfloor)}^{\hat{\alpha}^{LS}}(1-w)$. Dans ce cas, $X_{(\lfloor nw \rfloor)} \approx Q(w)$ et donc $F(x) \approx 1 - \hat{C}/x^{\hat{\alpha}^{LS}}$ pour $x \geq Q(w)$. Nous prenons $w = 90\%$.

Finalement, comme montré dans la section 3.4, la Value-at-Risk à un jour au niveau de confiance de 99% estimée à partir de la loi puissance est donnée par la formule 3.20 :

$$\hat{VaR}_a(t) = S_0 \left(1 - \exp \left(\frac{\hat{C}}{1 - a} \right)^{(\hat{\alpha}^{LS})^{-1}} \right).$$

La figure 3.2 représente le prix, la volatilité et la VaR à un jour et un niveau de confiance de 99% sur le cours de la BNP Paribas. Une dépendance générale de la VaR en fonction de la volatilité annuelle peut être observée sur cette figure.

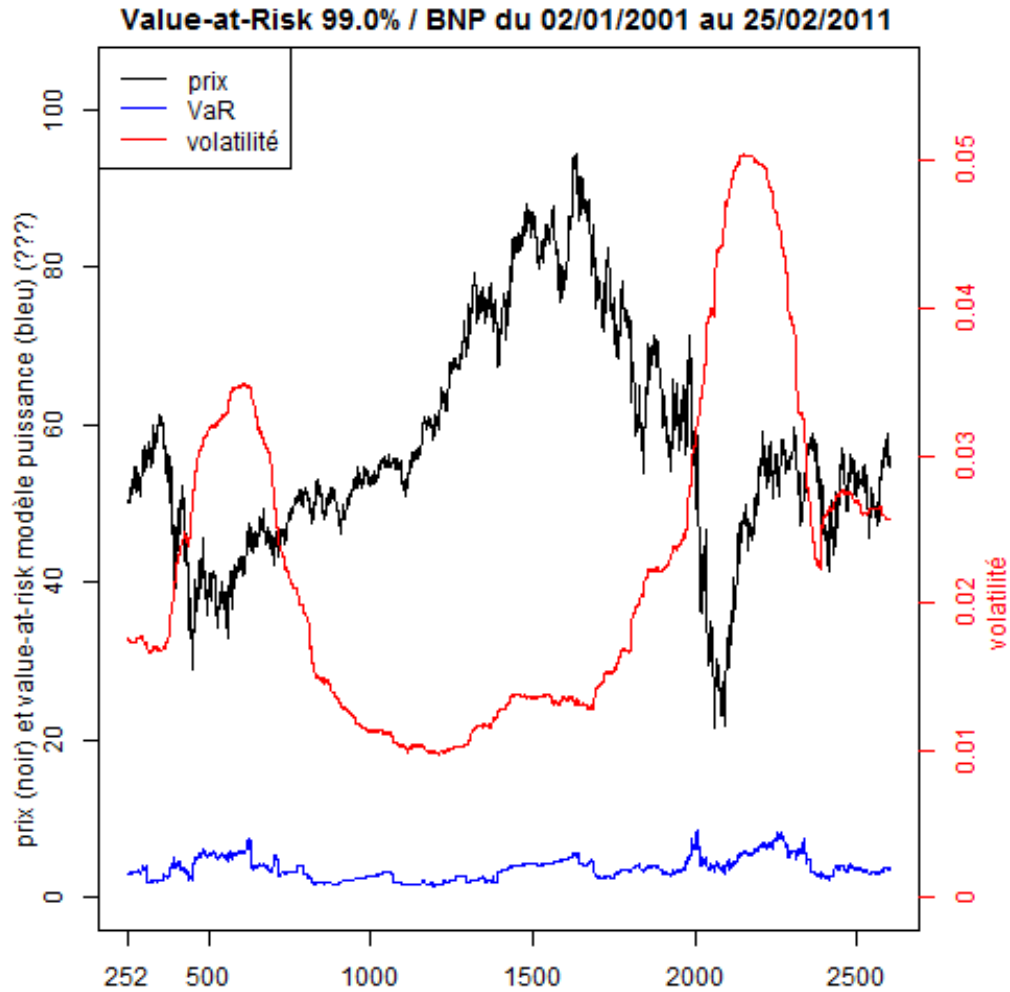


FIGURE 3.2 – VaR à 1% à 99% pour l'action BNP Paribas sur une fenêtre glissante de 252 jours avec les estimateurs $\hat{\alpha}^{LS}$ et \hat{C} .

Les paramètres de la loi puissance évalués sur une fenêtre glissante de 252 jours sont représentés dans la figure 3.3. Nous observons une très forte fluctuation des valeurs estimées de α et C au cours du temps. La méthode est assez instable et imprécise,

principalement à cause du faible nombre de données utilisé pour estimer ces paramètres. On observe néanmoins un phénomène de compensation entre les grandes valeurs de α et les petites valeurs de C (et inversement), qui conduit au final à une amplitude de fluctuations des valeurs de VaR plus réduite.

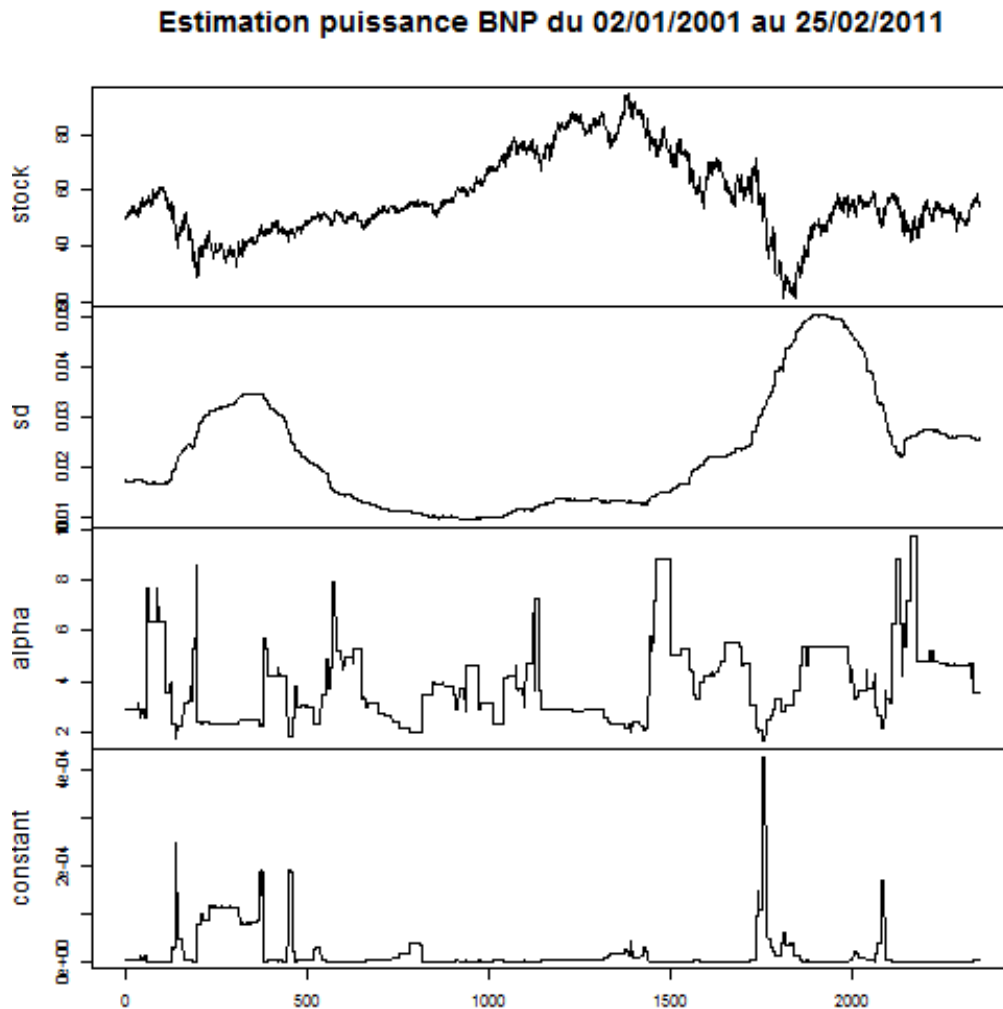


FIGURE 3.3 – Prix, écart-type, $\hat{\alpha}^{LS}$ et \hat{C} pour l'action BNP Paribas pour une fenêtre glissante de 252 jours

3.6 Backtesting sur la VaR

Le backtesting désigne un ensemble de techniques statistiques de validation permettant de vérifier l'adéquation des prévisions de la VaR. Il consiste à tester la pertinence de notre modélisation par les lois puissances dans le calcul de la Value-at-Risk. En effet, étant donné que la véritable valeur de la VaR n'est pas observable à posteriori, les techniques usuelles de validation des prévisions (*Mean Square Error*, *Mean Absolute*

Percentage Error...) ne peuvent pas être utilisées. Par ailleurs, il n'existe pas de proxy de la VaR auquel nous pouvons comparer nos prévisions. Pour cela, des procédures de backtesting ont été développées, cherchant toutes à tester l'hypothèse nulle selon laquelle la séquence de violations des estimations de la VaR est valide. D'où, en conclusion, le modèle d'évaluation du risque de marché est adéquat.

Comme chaque institution construit son propre modèle pour estimer la VaR, les procédures de validation sont généralement de type *model free* afin qu'elles puissent être appliquées à n'importe quel modèle interne. Ainsi, en utilisant les estimations de la VaR et les rendements observés de l'actif ou du portefeuille en général, on construit la séquence de dépassements suivante :

$$I_a(t) = \begin{cases} 1 & \text{si } S_{t+1} - S_t < \hat{V}aR_a(1), \\ 0 & \text{sinon.} \end{cases}$$

Cette séquence doit respecter deux conditions afin d'être validée :

1. L'hypothèse de couverture non conditionnelle : chaque jour la probabilité d'avoir un dépassement doit être égale au taux de couverture $(1 - a)$:

$$\mathbb{P}[I_a(t) = 1] = \mathbb{E}[I_a(t)] = a.$$

2. L'hypothèse d'indépendance : les dépassements de la séquence issus d'un même taux de couverture $(1 - a)$ doivent être indépendamment distribués. Autrement dit, ce n'est pas parce que l'on observe un dépassement aujourd'hui, qu'il y en aura un demain.

La littérature des méthodes de backtesting est large et variée. Cela a commencé par le backtesting binomial proposé par Kupiec [19] et qui cherche à déterminer la consistance de la fréquence de dépassements observés par rapport à la fréquence espérée par la modélisation de la VaR (le seuil de confiance). Ce test ignore la dépendance entre les dépassement et se concentre à vérifier la première condition. Chatfield [10] et Christoffersen [12] ont proposé des méthodes pour tester les deux hypothèses du processus de dépassements simultanément. Un test plus récent qui cherche à vérifier l'hypothèse d'indépendance a été suggéré par Christoffersen et Pelletier en 2004 [11]. L'idée est la suivante : si les dépassements de la VaR sont complètement indépendants alors les durées entre les dépassements doivent être indépendantes de la durée de temps qui s'est écoulé depuis le dernier dépassement. Néanmoins, les auteurs de ces méthodes testant les deux hypothèses, ont conclu qu'elles manquent d'efficacité pour détecter les mauvaises modélisations de la VaR pour des échantillons de petite taille. D'autres auteurs comme Berkowitz [7] ont développé une deuxième catégorie des tests *density forecast backtesting* se basant sur la densité des dépassements et non plus sur la fréquence. Nous renvoyons à [9] et [13] pour des études récapitulatives des différentes méthodes de backtesting.

Dans le cadre de ce rapport, vu le nombre limité de données, on opte pour le test de Kupiec [19]. Ainsi, la probabilité $\mathbb{P}[S_T - S_0 < \hat{V}aR_a(T)]$ est estimée par la fréquence de

dépassements de la VaR mesurée par les pertes postérieures, de la façon suivante : On considère les prix $(S_t)_{t=1,\dots,N}$ du cours et on pose $L_t = \log(S_{t+1}/S_t)$. On fixe la taille de la fenêtre n (égale à 252 jours dans ce chapitre). Pour $t = n, \dots, N$, on considère l'échantillon $\mathbb{L}_t = (L_{t-n}, \dots, L_{t-1})$ des log-rendements sur une fenêtre des n jours antérieurs. On calcule alors $\hat{\alpha}^{LS}$ et \hat{C} , les paramètres de la loi puissance modélisant \mathbb{L}_t , ainsi que $\hat{q}_{1-a}(t)$ son quantile de niveau $1 - a$. Pour $t = n + 1, \dots, N$, on compte le nombre moyen de fois où $L_t < \hat{q}_{1-a}(t)$ et on compare le résultat à $1 - a$.

On applique cette méthode aux log-rendements journaliers de l'action BNP Paribas entre 2001 et 2011 (même jeu de données que dans la section 3.5). Les résultats du back-testing sont présentés dans la figure 3.4. La VaR est calculée à un niveau de confiance de 99%, c'est-à-dire $a = 0.99$.

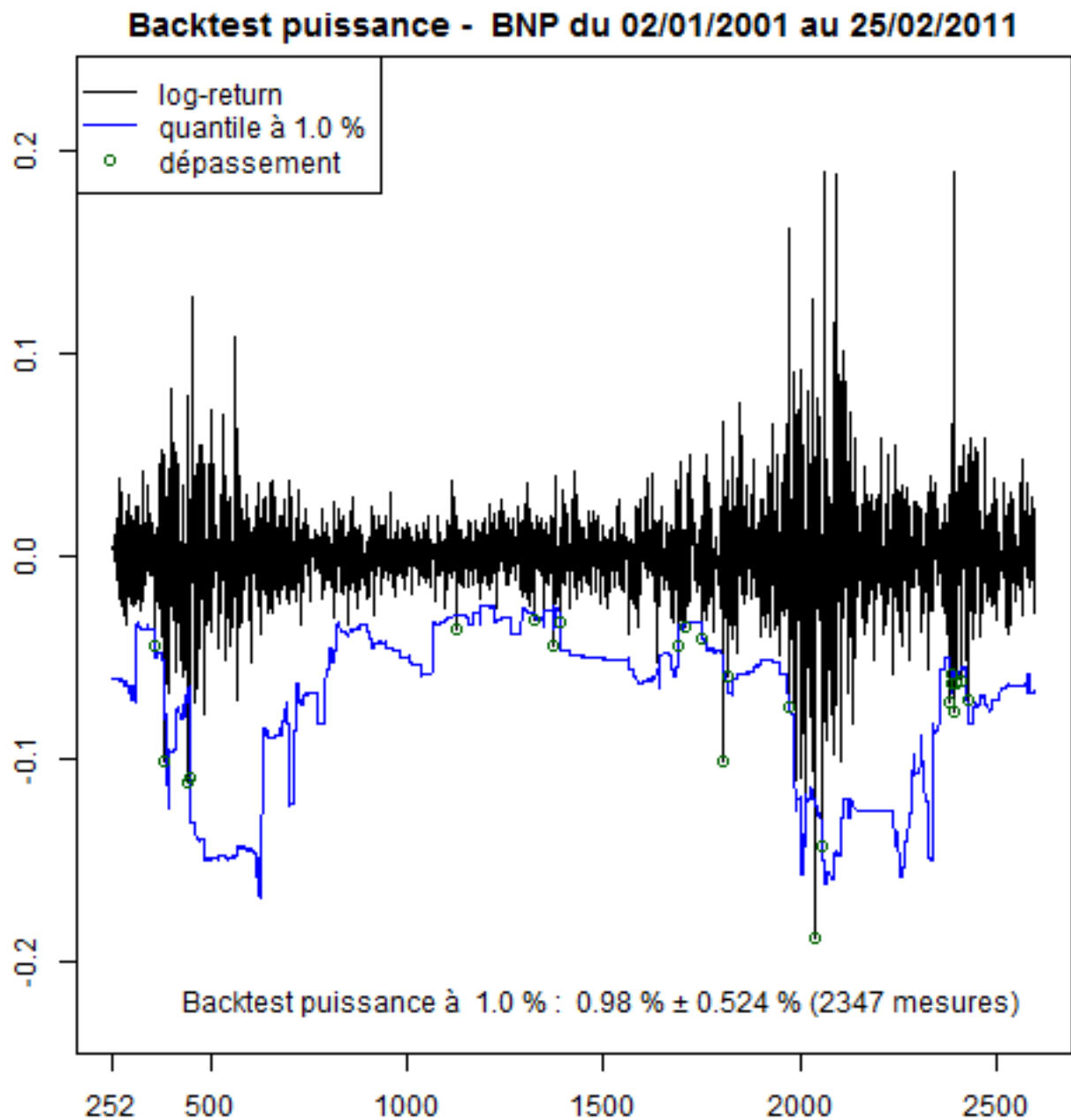


FIGURE 3.4 – Backtesting à 1% sur l’actif BNP Paribas sur une fenêtre glissante de 252 jours avec les estimateurs $\hat{\alpha}^{LS}$ et \hat{C} .

Chapitre 4

Analyse de sensibilité sur le modèle des lois puissances

Ce chapitre est consacré à l'analyse de sensibilité du calcul de la VaR par rapport aux variations de différentes entrées du modèle. Le but est de chercher les meilleures entrées possibles, comme la taille de la fenêtre de l'estimation, et de conclure sur la robustesse de nos estimateurs des paramètres de la loi puissance. On commence par tester la sensibilité à la taille de fenêtre de calcul de la VaR (prise par défaut dans le chapitre précédent à une année boursière, soit 252 jours). Ensuite, on teste notre méthode de calcul de la VaR avec des données bruitées et l'effet de ce bruitage sur nos estimateurs.

4.1 Sensibilité par rapport à la taille de la fenêtre

Dans l'étude numérique du chapitre précédent, on a opté pour le choix d'une fenêtre de 252 jours pour le calcul de la VaR. Cependant, ce choix est arbitraire et basé sur les habitudes des financiers. Ainsi, on essaiera dans cette section d'évaluer la VaR avec des différentes valeurs pour l'entrée taille de la fenêtre. On considère à chaque fois une taille fixe des fenêtres pour l'estimation des paramètres de la loi puissance, on calcule la VaR et on effectue un Back-testing pour comparer la performance de notre indicateur de risque suivant les différentes valeurs. Notre choix a été de travailler avec des fenêtres de 126 jours (une demi-année boursière), 252 jours (une année), 378 jours (une année et demi) et 504 jours (deux années).

Le tableau suivant résume les résultats du backtesting à 1 % sur les actifs BNP, CA, CNP, FGR, SAN, LI, RI, GLE et TEC pour les quatre valeurs de la taille de la fenêtre. Une première constatation, qu'on peut tirer de ce tableau, est que globalement la VaR estimée sur une fenêtre glissante de 126 jours donne en global un meilleur backtesting que celle estimée sur une fenêtre de 252 jours. En effet, les proportions de dépassement sont plus proches de 1 pour des fenêtres de taille 126 jours, tandis que les intervalles

de confiance sont semblables pour les deux cas. On explique ce résultat par le fait que estimer les paramètres de la loi puissance sur une fenêtre d'une demi-année boursière donne suffisamment de données pour obtenir de bons résultats tout en reculant moins dans le passé et en intégrant moins des crises anciennes dans l'estimation.

Titre	126 jours	252 jours	378 jours	504 jours
BNP	0.93 % \pm 0.497 %	0.98 % \pm 0.524 %	1.04 % \pm 0.553 %	1.00 % \pm 0.561 %
CA	1.01 % \pm 0.518 %	0.94 % \pm 0.512 %	0.90 % \pm 0.516 %	0.67 % \pm 0.459 %
CNP	1.05 % \pm 0.528 %	0.77 % \pm 0.464 %	0.81 % \pm 0.490 %	0.67 % \pm 0.459 %
FGR	1.09 % \pm 0.538 %	0.77 % \pm 0.464 %	0.76 % \pm 0.476 %	0.76 % \pm 0.490 %
SAN	1.09 % \pm 0.538 %	0.68 % \pm 0.438 %	0.72 % \pm 0.462 %	0.90 % \pm 0.534 %
LI	1.13 % \pm 0.548 %	0.81 % \pm 0.477 %	0.72 % \pm 0.462 %	0.62 % \pm 0.442 %
RI	0.77 % \pm 0.452 %	0.72 % \pm 0.451 %	0.67 % \pm 0.448 %	0.76 % \pm 0.490 %
GLE	0.85 % \pm 0.475 %	0.85 % \pm 0.489 %	0.81 % \pm 0.490 %	0.86 % \pm 0.520 %
TEC	1.05 % \pm 0.528 %	1.07 % \pm 0.546 %	0.90 % \pm 0.516 %	0.76 % \pm 0.490 %

TABLE 4.1 – Backtesting sur les actions BNP, CA, CNP, FGR, SAN, LI, RI, GLE et TEC sur une fenêtre glissante de 126 jours, 252 jours, 378 jours et 504 jours.

La figure 4.1 montre l'effet du changement de taille de la fenêtre sur nos estimateurs de la loi puissance. En terme de volatilité, plus on augmente la taille de la fenêtre, plus les deux pics de la volatilité s'étalent dans le temps. Par ailleurs, les courbes des estimateurs de α et de C possèdent de moins en moins de fluctuation en augmentant la taille de la fenêtre. En effet, avec des fenêtres de calcul plus grandes, on utilise plus de données, donc plus d'événements passés agissent sur notre estimation. Ainsi, une pic dans les prix de l'actif agira plus longtemps sur la volatilité, d'où le phénomène d'étalement. De même, ce recul dans le passé assure que les estimations de α et C deviennent plus stable, avec moins de fluctuation. En conséquence, pour une taille de fenêtre de 126 jours, même si ce choix donne un meilleur backtesting, il représente des estimations très instables avec beaucoup de fluctuations. Pour ces raisons, on continue dans ce qui suit à estimer avec des fenêtres de 252 jours.

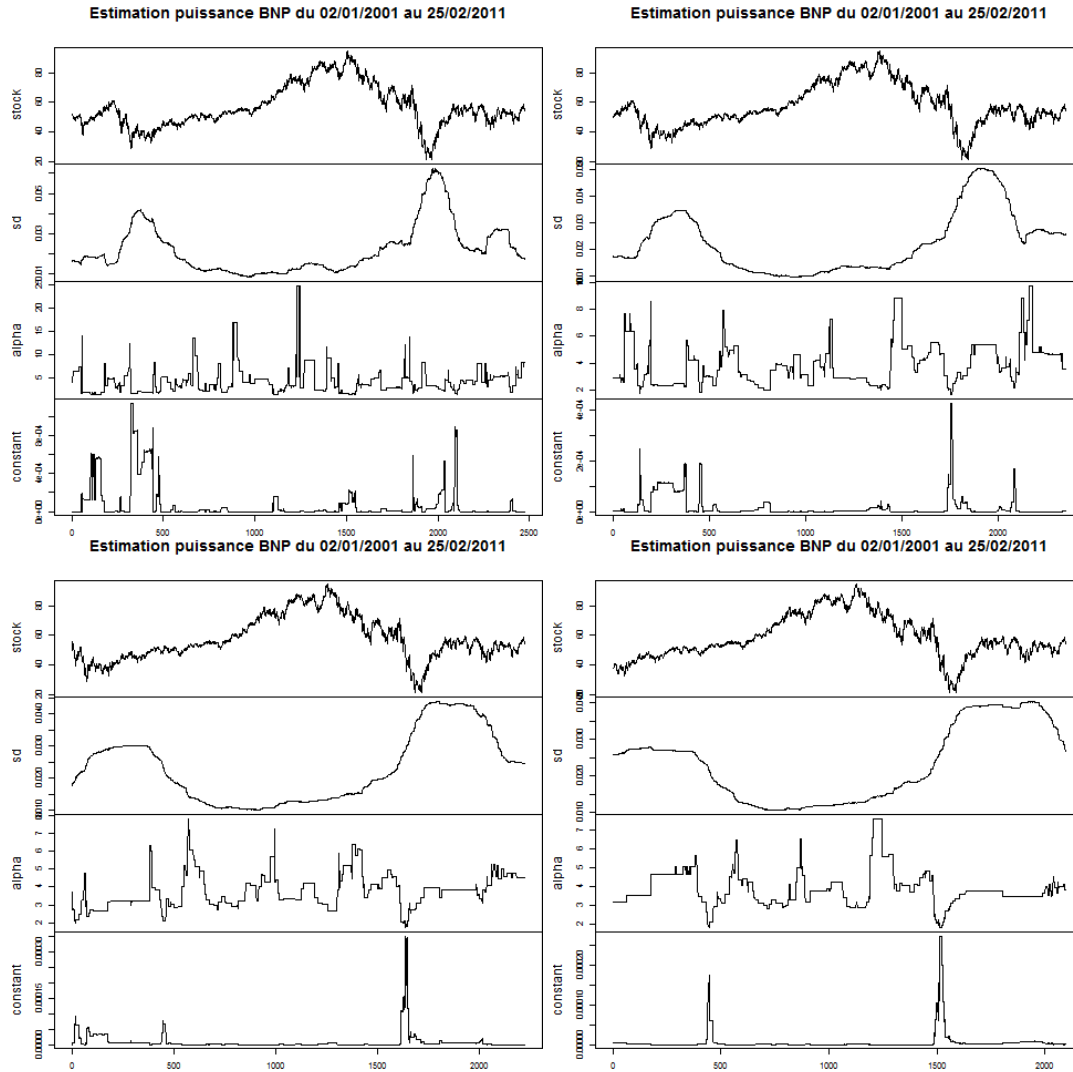


FIGURE 4.1 – Les paramètres de la loi puissances calculés respectivement de gauche à droite pour une fenêtre glissante de 126, 252, 378 et 504 jours.

4.2 Sensibilité par rapport aux données bruitées

Dans cette section, on teste la robustesse de notre méthode d'estimation. On cherche alors à étudier la capacité de nos estimateurs de ne pas être modifiés par une petite perturbation des données. On considère comme données d'entrée le prix de fermeture des actifs et on le perturbe par l'ajout d'un bruit gaussien de variance $1/10$ de sa variance.

Dans la figure 4.2, nous représentons la Value-at-Risk à un jour au niveau de risque de 1% estimée à la fois à partir des données réelles et à la fois à partir des données bruitées du prix de fermeture de l'actif BNP. La figure 4.3 montre l'effet de l'ajout de

bruit par les courbes de différence entre la VaR et la volatilité estimée à partir des données réelles et celles estimée à partir des données bruitées. On constate que la VaR des données bruitées est souvent au dessus de celle des données réelles. Cela s'explique par le fait que les données bruitées sont plus volatiles en raison de l'ajout d'un bruit normal, donc elles auront une VaR plus élevée. Le back-testing à 1% sur une fenêtre glissante de 252 jours est donné par la figure 4.5 pour l'action BNP Paribas, ainsi que par le tableau 4.2 résumant les résultat du back-testing pour les action BNP, CA, CNP, FGR, GLE, LI, RI, SAN et TEC.

Les différentes estimations des paramètres de loi puissance se trouvent dans la figure 4.4. On remarque une ressemblance générale dans les formes des courbes des paramètres. La volatilité garde la même allure avec une courbe plus élevée. Cela s'explique par le fait que l'ajout d'un bruit normal rend l'actif plus volatile. De même façon, on remarque un comportement semblable pour les estimateurs de α et de la constante de Weissman : même allure avec une augmentation.

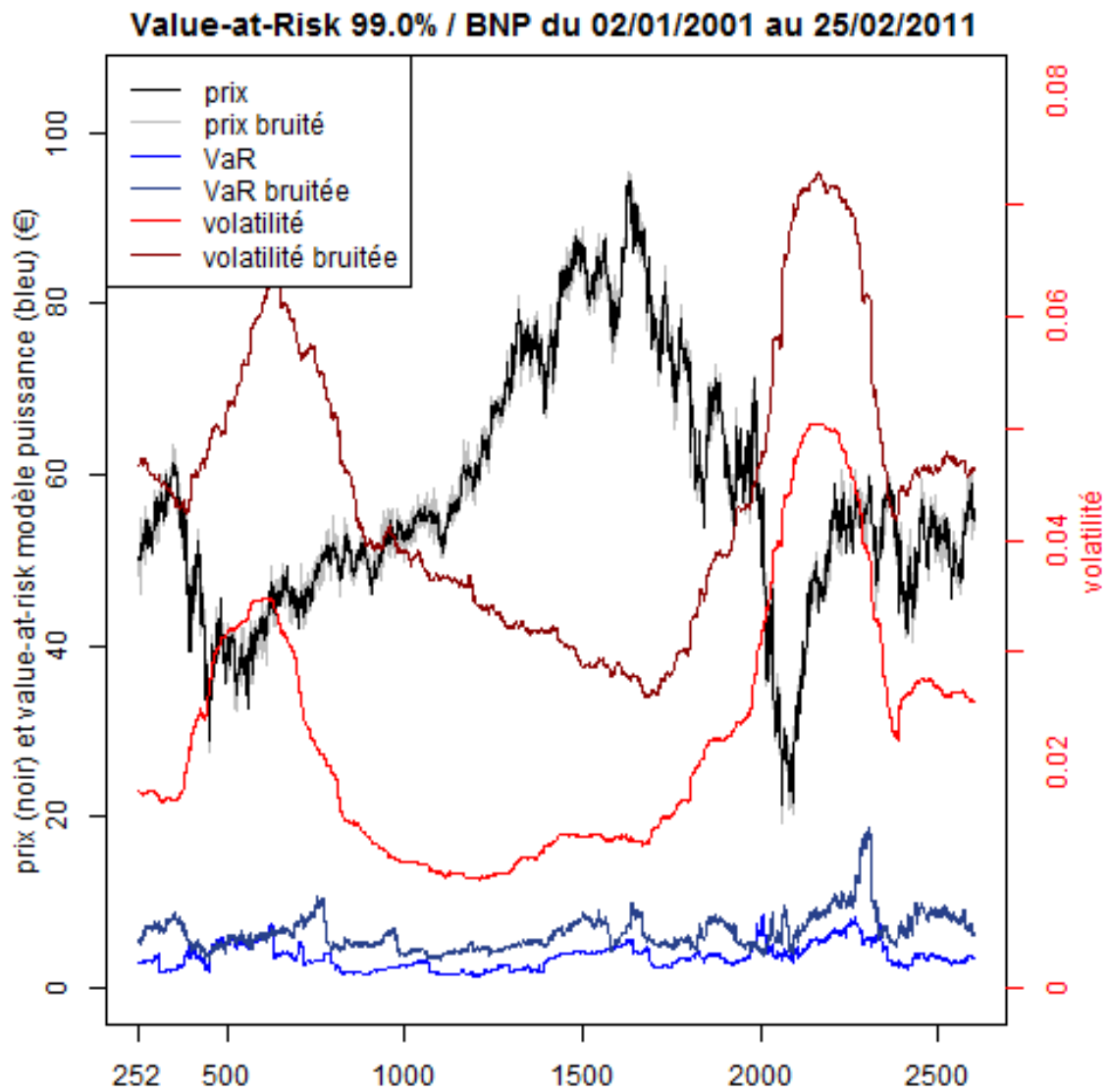


FIGURE 4.2 – VaR à 1 jour à 99% pour l'action BNP Paribas estimée une fois à partir des données réelles et une fois à partir des données bruitées.

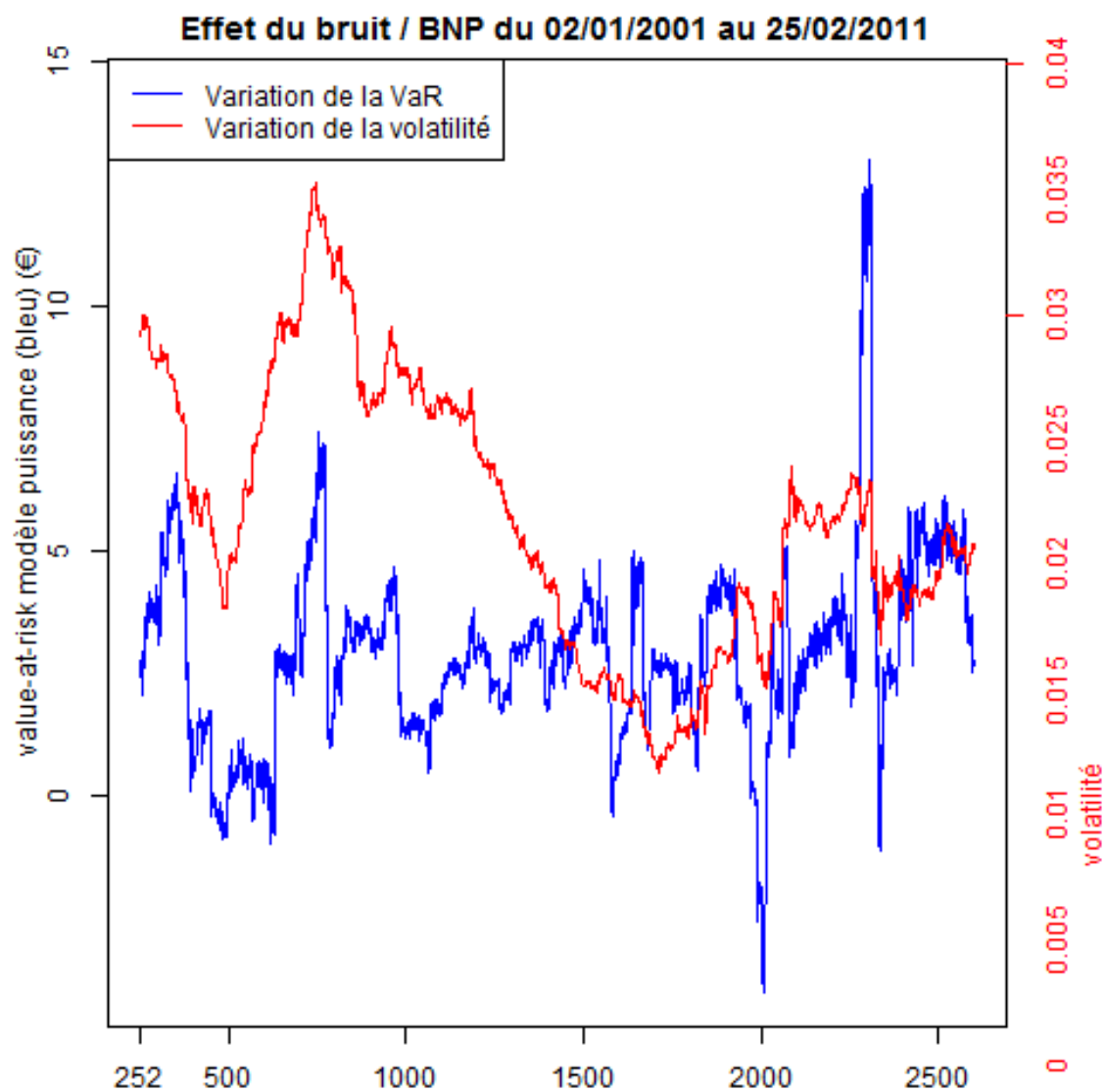


FIGURE 4.3 – Effet du bruit sur la VaR et sur la volatilité pour l'action BNP Paribas.

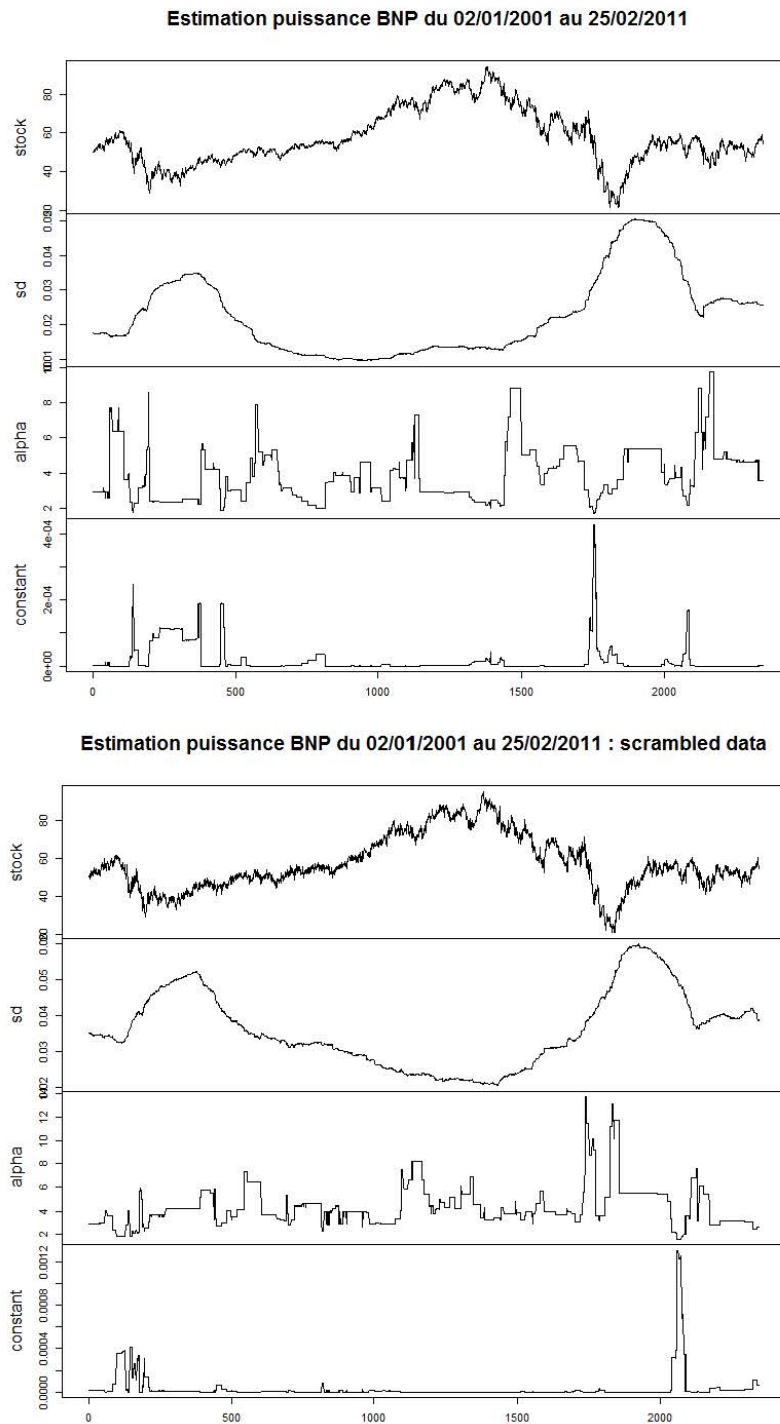


FIGURE 4.4 – Prix, écart-type, $\hat{\alpha}$ et \hat{C} pour l'action BNP Paribas pour une fenêtre glissante de 252 jours.

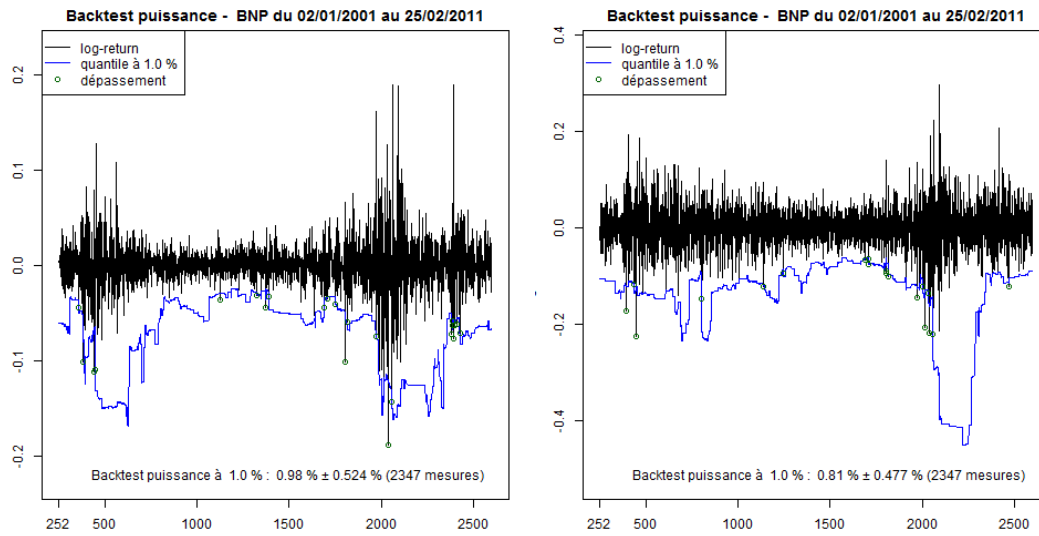


FIGURE 4.5 – Backtesting pour données réelles (gauche) et données bruitées (droite) de l'action BNP Paribas pour une fenêtre glissante de 252 jours.

Titre	Données réelles	Données bruitées
BNP	0.98 % \pm 0.524 %	0.81 % \pm 0.477 %
CA	0.937 % \pm 0.512 %	0.554 % \pm 0.395 %
CNP	0.767 % \pm 0.464 %	0.767 % \pm 0.464 %
FGR	0.767 % \pm 0.464 %	0.724 % \pm 0.451 %
SAN	0.682 % \pm 0.438 %	1.02 % \pm 0.535 %
LI	0.81 % \pm 0.477 %	0.682 % \pm 0.438 %
RI	0.724 % \pm 0.451 %	0.639 % \pm 0.424 %
GLE	0.852 % \pm 0.489 %	0.81 % \pm 0.477 %
TEC	1.07 % \pm 0.546 %	1.15 % \pm 0.567 %

TABLE 4.2 – Back-testing pour données réelles et données bruitées des actifs BNP, CA, CNP, FGR, SAN, LI, RI et LI.

Conclusion

Dans ce chapitre, on a testé la sensibilité du modèle puissance par rapport à certains éléments soit le changement des paramètres d'entrée comme la taille de la fenêtre, soit les perturbations extérieures. Compte tenu du nombre réduit des données, les résultats ne sont pas concluants. Cependant, cela nous a permis d'identifier certaines propriétés. En effet, l'augmentation de la taille de la fenêtre conduit à des estimateurs avec moins de fluctuations. En prenant aussi en considération les résultats du backtesting, où le test est validé que pour les tailles de 126 et 252 jours, nous continuons à travailler avec des fenêtres glissantes de 252 jours. D'autre part, nous constatons que nos estimateurs présentent une sorte de stabilité par rapport à l'ajout de bruit en gardant la même allure de courbe avec un changement d'échelle. Ces informations nous ont encouragé à continuer avec le modèle puissance et à l'améliorer par d'autres approches. C'est ainsi que l'on va introduire un nouveau critère de crise et qu'on va classer nos données selon ce critère par l'application du modèle de Markov caché. Le chapitre suivant représente le modèle de Markov caché, ses problématiques canoniques et les différents algorithmes de résolution. Ensuite, dans le chapitre 6, nous montrons l'application sur nos données et définissons une nouvelle méthode de calcul de la VaR intégrant le critère crise.

Chapitre 5

Modèle de Markov Caché

Les bases des modèles de Markov cachés ont été publiées dans une série d'articles à la fin des années 1960 et au début des années 1970, autour de Leonard E. Baum [1–5]. Ces modèles supposent, tout en disposant d'une série d'observations, l'existence d'une variable cachée qui est à l'origine de ces observations, et cherche à expliciter ce caractère caché. Ce chapitre se divise en trois sections. Nous commençons par présenter la modélisation et les problèmes canoniques liés au modèle de Markov caché. Ensuite, on résout l'un de ces problèmes par l'introduction de l'algorithme de Baum-Welsh [14, 22] qui sert à estimer les paramètres du modèle caché à partir d'une série d'observations. Une dernière section est consacrée à l'algorithme de Viterbi et à la reproduction de la chaîne cachée.

5.1 Problématique

On considère une chaîne de Markov $X = (X_n)_{n \geq 0}$ à valeurs dans un espace d'état \mathcal{X} fini. On note ρ sa loi initiale et $Q = [Q(x, x')]_{x, x' \in \mathcal{X}}$ sa matrice de transition. La chaîne X n'est pas directement observable, elle est dite cachée. On dispose d'une suite d'observations $Y = (Y_n)_{n \geq 0}$ à valeurs dans un espace d'observations \mathcal{Y} .

Notation : On désigne par $X_{k:l}$ la séquence (X_k, \dots, X_l) si $k \leq l$ et l'ensemble vide si $k > l$.

Hypothèse 5.1 (Canal d'observation sans mémoire). *Pour tout $n \geq 1$, $x_{0:n} \in \mathcal{X}^{n+1}$ et $y_{0:n} \in \mathcal{Y}^{n+1}$:*

$$\mathbb{P}(Y_{0:n} = y_{0:n} \mid X_{0:n} = x_{0:n}) = \prod_{k=0}^n \mathbb{P}(Y_k = y_k \mid X_k = x_k).$$

Cette hypothèse caractérise la façon dont la loi des observations dépend de celle du processus d'état. En d'autres termes, conditionnellement aux états X_k , les observations Y_k sont mutuellement indépendantes et chaque observation ne dépend que de l'état X_k . On notera $\psi(x, y)$ la vraisemblance associée, que l'on appelle la matrice d'émission.

Définition 5.1. Une matrice $\psi = [\psi(x, y)]_{x \in \mathcal{X}, y \in \mathcal{Y}}$ est appelée matrice de Markov sur $\mathcal{X} \times \mathcal{Y}$ lorsque :

(i) $\psi(x, y) \geq 0$ pour tout $x \in \mathcal{X}$, $y \in \mathcal{Y}$,

(ii) $\sum_{y \in \mathcal{Y}} \psi(x, y) = 1$ pour tout $x \in \mathcal{X}$.

C'est donc une famille de lois de probabilités sur \mathcal{Y} , indexée par les éléments de \mathcal{X} .

On vérifie aisément que la matrice d'émission ψ définie par :

$$\psi(x, y) = \mathbb{P}(Y_k = y | X_k = x), \quad x \in \mathcal{X}, \quad y \in \mathcal{Y},$$

est de Markov. On suppose donc implicitement que la loi de Y_k sachant X_k ne dépend pas de l'instant k .

On en déduit le résultat suivant.

Proposition 5.1. La loi jointe du couple (X, Y) est entièrement caractérisée par la loi initiale ρ , la matrice de transition Q de la chaîne de Markov X et la matrice d'émission ψ de la manière suivante :

$$\mathbb{P}(X_{0:n} = x_{0:n}, Y_{0:n} = y_{0:n}) = \rho(x_0) \psi(x_0, y_0) \prod_{k=1}^n [Q(x_{k-1}, x_k) \psi(x_k, y_k)], \quad (5.1)$$

pour tout $n \in \mathbb{N}$, $x_{0:n} \in \mathcal{X}^{n+1}$ et $y_{0:n} \in \mathcal{Y}^{n+1}$. En particulier le couple (X, Y) est markovien.

Ainsi, un modèle de Markov caché (X, Y) , noté généralement (MMC), est entièrement caractérisé par :

(i) la loi initiale ρ du processus d'état X :

$$\rho(x) = \mathbb{P}(X_0 = x), \quad x \in \mathcal{X};$$

(ii) la matrice de transition Q du processus d'état X :

$$Q(x, x') = \mathbb{P}(X_{k+1} = x' | X_k = x), \quad x, x' \in \mathcal{X};$$

(iii) la matrice d'émission du processus Y sachant le processus X :

$$\psi(x, y) = \mathbb{P}(Y_k = y | X_k = x), \quad x \in \mathcal{X}, \quad y \in \mathcal{Y}.$$

On désigne par $M = (\rho, Q, \psi)$ les paramètres du modèle de Markov caché.

Les problématiques du modèle de Markov caché peuvent être résumées dans les trois problèmes canoniques suivants :

1. **Problème d'évaluation** : étant donnés les paramètres $M = (\rho, Q, \psi)$ du modèle, calculer la vraisemblance d'une séquence d'observations $Y_{0:n}$, c'est-à-dire la distribution de probabilité de la séquence en fonctions des paramètres du modèle M .

2. **Problème de décodage ou de reconnaissance** : étant donnés les paramètres $M = (\rho, Q, \psi)$ du modèle et une séquence d'observations $Y_{0:n}$, déterminer la séquence d'états $X_{0:n}$ qui est le plus vraisemblablement à l'origine de ces observations $Y_{0:n}$.
3. **Problème de ré-estimation ou d'apprentissage** : étant donnée une séquence d'observation $Y_{0:n}$, déterminer les paramètres $M = (\rho, Q, \psi)$ du modèle qui maximise la vraisemblance d'apparition de cette séquence d'observation.

5.2 Algorithme de Baum-Welsh et problème de ré-estimation

De part le caractère caché des états, on peut retrouver des situations où les paramètres du modèle sont eux aussi à estimer. Un algorithme particulièrement efficace dans l'estimation de ces paramètres est l'algorithme de Baum-Welsh. Cet algorithme représente une procédure de maximisation locale. Il s'agit d'une dérivé de l'algorithme Expectation-Maximization (EM) où on cherche à ajuster les paramètres $M = (\rho, Q, \psi)$ au critère d'optimisation "mieux correspondre aux observations". En d'autres termes, on cherche à déterminer $M = (\rho, Q, \psi)$ qui maximise localement la fonction de vraisemblance :

$$M \mapsto \mathcal{L}_k(M) = \mathbb{P}_M(Y_{0:n} = y_{0:n})$$

où \mathbb{P}_M est la probabilité sous laquelle le modèle M est valide et $y_{0:n}$ est un vecteur de \mathcal{Y}^{n+1} .

L'algorithme de Baum-Welsh fait appel à l'algorithme "forward" de Baum et à l'algorithme "backward" de Baum qu'on va présenter dans les sections qui suivent. Ces deux algorithmes cherchent à calculer efficacement la distribution de probabilité de la séquence d'observation $Y_{0:n}$ en fonction des paramètres du modèle.

5.2.1 Algorithme "forward" de Baum

On définit la loi jointe de $Y_{0:n}$ et X_n :

$$\alpha_x^n(y_{0:n}) = \mathbb{P}(Y_{0:n} = y_{0:n}, X_n = x) \quad (5.2)$$

pour $n \in \mathbb{N}$ et $y_{0:n} \in \mathcal{Y}^{n+1}$, $x \in \mathcal{X}$, ainsi que la variable "forward" :

$$p_n(x) = \alpha_x^n(Y_{0:n}) \quad (5.3)$$

pour $n \in \mathbb{N}$ et $x \in \mathcal{X}$.

Théorème 5.1 (Algorithme "forward" de Baum). *La suite $(p_n)_{n \in \mathbb{N}}$ des variables forward est solution de l'équation récurrente suivante :*

$$p_{n+1}(x) \leftarrow \psi(x, Y_{n+1}) \sum_{x' \in \mathcal{X}} Q(x', x) p_n(x'), \quad n \geq 0, \quad x \in \mathcal{X}, \quad (5.4)$$

$$p_0(x) \leftarrow \rho(x) \psi(x, Y_0), \quad x \in \mathcal{X}. \quad (5.5)$$

Preuve. Pour $n = 0$, on montre aisément que

$$\alpha_x^0(y) = \mathbb{P}(X_0 = x, Y_0 = y) = \mathbb{P}(Y_0 = y | X_0 = x) \mathbb{P}(X_0 = x) = \psi(x, y) \rho(x).$$

Soit $n \geq 1$, on a

$$\begin{aligned} \alpha_{x_n}^n(y_{0:n}) &= \mathbb{P}(Y_{0:n} = y_{0:n}, X_n = x_n) \\ &= \sum_{x_{0:n-1} \in \mathcal{X}^n} \mathbb{P}(Y_{0:n} = y_{0:n}, X_{0:n-1} = x_{0:n-1}, X_n = x_n) \\ &= \sum_{x_{0:n-1} \in \mathcal{X}^n} \rho(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \psi(x_0, y_0) \dots \psi(x_n, y_n). \end{aligned}$$

Nous obtenons ainsi

$$\begin{aligned} \alpha_{x_{n+1}}^{n+1}(y_{0:n+1}) &= \\ &= \sum_{x_{0:n} \in \mathcal{X}^{n+1}} \rho(x_0) Q(x_0, x_1) \dots Q(x_n, x_{n+1}) \psi(x_0, y_0) \dots \psi(x_{n+1}, y_{n+1}) \\ &= \psi(x_{n+1}, y_{n+1}) \sum_{x_n \in \mathcal{X}} Q(x_n, x_{n+1}) \sum_{x_{0:n-1} \in \mathcal{X}^n} \rho(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \times \\ &\quad \times \psi(x_0, y_0) \dots \psi(x_n, y_n) \\ &= \psi(x_{n+1}, y_{n+1}) \sum_{x_n \in \mathcal{X}} Q(x_n, x_{n+1}) \alpha_{x_n}^n(y_{0:n}). \end{aligned}$$

Le résultat découle de cette dernière équation en remplaçant $y_{0:n+1}$ par $Y_{0:n+1}$. □

La loi des observations $Y_{0:n}$ peut être calculée de la façon suivante :

$$\mathbb{P}(Y_{0:n} = y_{0:n}) = \sum_{x \in \mathcal{X}} \alpha_x^n(y_{0:n}), \quad (5.6)$$

pour tout $n \in \mathbb{N}$ et $y_{0:n} \in \mathcal{Y}^{n+1}$.

5.2.2 Algorithme "backward" de Baum

Soit $n \geq 1$ un instant fixé. Pour tout instant $k < n$, on introduit la loi conditionnelle des observations $Y_{k+1:n}$ sachant que $X_k = x$:

$$\beta_x^k(y_{k+1:n}) = \mathbb{P}(Y_{k+1:n} = y_{k+1:n} | X_k = x) \quad (5.7)$$

pour $x \in \mathcal{X}$, $y_{k+1:n} \in \mathcal{Y}^{n-k}$, ainsi que la variable "backward" :

$$v_k(x) = \beta_x^k(Y_{k+1:n}) \quad (5.8)$$

pour $x \in \mathcal{X}$.

Remarque 5.1. Rappelons que, conditionnellement à l'événement $\{X_k = x\}$, la suite $(X_l)_{l>k}$ est une chaîne de Markov de loi initiale :

$$\mathbb{P}(X_{k+1} = x' | X_k = x) = Q(x, x'), \quad x' \in \mathcal{X}$$

et de matrice de transition Q .

Théorème 5.2 (Algorithme "backward" de Baum). La séquence $v_{1:n}$ des variables "backward" est solution de l'équation récurrente rétrograde suivante :

$$v_k(x) \leftarrow \sum_{x' \in \mathcal{X}} Q(x, x') \psi(x', Y_{k+1}) v_{k+1}(x'), \quad 0 \leq k < n, \quad x \in \mathcal{X}, \quad (5.9)$$

$$v_n(x) \leftarrow 1, \quad x \in \mathcal{X}. \quad (5.10)$$

Preuve. Notons d'abord que pour $k = n - 1$, on a :

$$\begin{aligned} \beta_x^{n-1}(y) &= \mathbb{P}(Y_n = y | X_{n-1} = x) \\ &= \sum_{x' \in \mathcal{X}} \mathbb{P}(Y_n = y, X_n = x' | X_{n-1} = x) \\ &= \sum_{x' \in \mathcal{X}} \mathbb{P}(Y_n = y | X_n = x', X_{n-1} = x) \mathbb{P}(X_n = x' | X_{n-1} = x) \\ &= \sum_{x' \in \mathcal{X}} \mathbb{P}(Y_n = y | X_n = x') \mathbb{P}(X_n = x' | X_{n-1} = x) \\ &= \sum_{x' \in \mathcal{X}} \psi(x', y) Q(x, x'). \end{aligned}$$

Ainsi en utilisant (5.10), on obtient (5.9) pour $k = n - 1$. D'autre part, il découle de la Remarque 5.1 et de la Proposition 5.1 que :

$$\begin{aligned} \beta_{x_k}^k(y_{k+1:n}) &= \mathbb{P}(Y_{k+1:n} = y_{k+1:n} | X_k = x_k) \\ &= \sum_{x_{k+1:n} \in \mathcal{X}^{n-k}} \mathbb{P}(Y_{k+1:n} = y_{k+1:n}, X_{k+1:n} = x_{k+1:n} | X_k = x_k) \\ &= \sum_{x_{k+1:n} \in \mathcal{X}^{n-k}} Q(x_k, x_{k+1}) \dots Q(x_{n-1}, x_n) \psi(x_{k+1}, y_{k+1}) \dots \psi(x_n, y_n), \end{aligned}$$

pour tout $x_k \in \mathcal{X}$, $y_{k+1:n} \in \mathcal{Y}^{n-k}$.

Ainsi

$$\begin{aligned} \beta_{x_k}^k(y_{k+1:n}) &= \sum_{x_{k+1:n} \in \mathcal{X}^{n-k}} Q(x_k, x_{k+1}) \dots Q(x_{n-1}, x_n) \psi(x_{k+1}, y_{k+1}) \dots \psi(x_n, y_n) \\ &= \sum_{x_{k+1} \in \mathcal{X}} Q(x_k, x_{k+1}) \psi(x_{k+1}, y_{k+1}) \sum_{x_{k+2:n} \in \mathcal{X}^{n-k-1}} Q(x_{k+1}, x_{k+2}) \dots Q(x_{n-1}, x_n) \\ &\quad \times \psi(x_{k+2}, y_{k+2}) \dots \psi(x_n, y_n) \\ &= \sum_{x_{k+1} \in \mathcal{X}} Q(x_k, x_{k+1}) \psi(x_{k+1}, y_{k+1}) \beta_{x_{k+1}}^{k+1}(y_{k+2:n}) \end{aligned}$$

d'où le résultat. □

On peut conclure que les deux algorithmes "forward/backward" de Baum permettent de calculer les quantités $\alpha_x^n(y_{0:n})$ et $\beta_x^k(y_{k+1:n})$ d'une manière récurrente non-coûteuse, facile à implémenter.

5.2.3 L'algorithme de Baum-Welsh

Considérons une séquence d'observations $Y_{0:n}$ et un modèle initial M . La méthode de ré-estimation de Baum-Welsh présente une procédure de maximisation locale qui cherche à ré-estimer d'une manière itérative les paramètres du modèle M jusqu'à arriver à un maximum local en termes de mieux correspondre aux observations $Y_{0:n}$.

Considérons la probabilité de transition $x \rightarrow x'$ sachant le modèle M et les observations $Y_{0:n} = y_{0:n}$:

$$\xi_{x,x'}^k(y_{0:n}) = \mathbb{P}_M(X_k = x, X_{k+1} = x' | Y_{0:n} = y_{0:n}).$$

Ce dernier terme peut s'écrire en fonction de $\alpha_x^k(y_{0:k})$ et $\beta_x^k(y_{k+1:n})$:

$$\begin{aligned} \xi_{x,x'}^k(y_{0:n}) &= \frac{\mathbb{P}_M(X_k = x, X_{k+1} = x', Y_{0:n} = y_{0:n})}{\mathbb{P}_M(Y_{0:n} = y_{0:n})} \\ &= \frac{\mathbb{P}_M(X_k = x, X_{k+1} = x', Y_{0:n} = y_{0:n})}{\sum_{\bar{x}, \bar{x}' \in \mathcal{X}} \mathbb{P}_M(X_k = \bar{x}, X_{k+1} = \bar{x}' | Y_{0:n} = y_{0:n})} \\ &= \frac{\alpha_x^k(y_{0:k}) Q(x, x') \psi(x', y_{k+1}) \beta_{x'}^{k+1}(y_{k+2:n})}{\sum_{\bar{x}, \bar{x}' \in \mathcal{X}} \alpha_{\bar{x}}^k(y_{0:k}) Q(\bar{x}, \bar{x}') \psi(\bar{x}', y_{k+1}) \beta_{\bar{x}'}^{k+1}(y_{k+2:n})}. \end{aligned}$$

En effet,

$$\begin{aligned} &\mathbb{P}_M(X_k = x_k, X_{k+1} = x_{k+1}, Y_{0:n} = y_{0:n}) \\ &= \sum_{\substack{x_{0:k-1} \in \mathcal{X}^k \\ x_{k+2:n} \in \mathcal{X}^{n-k-1}}} \mathbb{P}_M(X_{0:n} = x_{0:n}, Y_{0:n} = y_{0:n}) \\ &= \sum_{\substack{x_{0:k-1} \in \mathcal{X}^k \\ x_{k+2:n} \in \mathcal{X}^{n-k-1}}} \rho(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \psi(x_0, y_0) \dots \psi(x_n, y_n) \\ &= \left[\sum_{x_{0:k-1} \in \mathcal{X}^k} \rho(x_0) Q(x_0, x_1) \dots Q(x_{k-1}, x_k) \psi(x_0, y_0) \dots \psi(x_k, y_k) \right] \times Q(x_k, x_{k+1}) \times \\ &\quad \psi(x_{k+1}, y_{k+1}) \times \left[\sum_{x_{k+2:n} \in \mathcal{X}^{n-k-1}} Q(x_{k+1}, x_{k+2}) \dots Q(x_{n-1}, x_n) \psi(x_{k+2}, y_{k+2}) \dots \psi(x_n, y_n) \right] \\ &= \alpha_{x_k}^k(y_{0:k}) Q(x_k, x_{k+1}) \psi(x_{k+1}, y_{k+1}) \beta_{x_{k+1}}^{k+1}(y_{k+2:n}). \end{aligned}$$

De même, on peut obtenir

$$\mathbb{P}_M(X_k = x, Y_{0:n} = y_{0:n}) = \alpha_x^k(y_{0:k})\beta_x^k(y_{k+1:n}). \quad (5.11)$$

On pose :

$$\gamma_x^k(y_{0:n}) = \sum_{x' \in \mathcal{X}} \xi_{x,x'}^k(y_{0:n}).$$

Remarquons que $\gamma_x^k(y_{0:n}) = \mathbb{P}_M(X_k = x | Y_{0:n} = y_{0:n})$. Ainsi, l'expression (5.11) nous donne :

$$\gamma_x^k(y_{0:n}) = \frac{\alpha_x^k(y_{0:k})\beta_x^k(y_{k+1:n})}{\sum_{x' \in \mathcal{X}} \alpha_{x'}^k(y_{0:k})\beta_{x'}^k(y_{k+1:n})}.$$

Une sommation en k de 0 à $n-1$ de cette expression donne :

$$\begin{aligned} \sum_{k=0}^{n-1} \gamma_x^k(y_{0:n}) &= \sum_{k=0}^{n-1} \mathbb{P}_M(X_k = x | Y_{0:n} = y_{0:n}) \\ &= \sum_{k=0}^{n-1} \mathbb{E}_M(\mathbf{1}_{X_k=x} | Y_{0:n} = y_{0:n}) \\ &= \mathbb{E}_M \left(\sum_{k=0}^{n-1} \mathbf{1}_{X_k=x} | Y_{0:n} = y_{0:n} \right) \\ &= \mathbb{E}_M(\#\{0 \leq k < n : X_k = x\} | Y_{0:n} = y_{0:n}). \end{aligned}$$

Cette somme peut être interprétée comme l'estimateur du nombre de transitions issues de x de la séquence $X_{0:n}$ au vu des observations $Y_{0:n} = y_{0:n}$.

De même :

$$\begin{aligned} \sum_{k=0}^{n-1} \xi_{x,x'}^k(y_{0:n}) &= \sum_{k=0}^{n-1} \mathbb{P}_M(X_k = x, X_{k+1} = x' | Y_{0:n} = y_{0:n}) \\ &= \sum_{k=0}^{n-1} \mathbb{E}_M(\mathbf{1}_{\{X_k=x, X_{k+1}=x'\}} | Y_{0:n} = y_{0:n}) \\ &= \mathbb{E}_M \left(\sum_{k=0}^{n-1} \mathbf{1}_{\{X_k=x, X_{k+1}=x'\}} \middle| Y_{0:n} = y_{0:n} \right) \\ &= \mathbb{E}_M(\#\{0 \leq k < n : X_k = x, X_{k+1} = x'\} | Y_{0:n} = y_{0:n}) \end{aligned}$$

représente l'estimateur du nombre de transitions $x \rightarrow x'$ de la séquence $X_{0:n}$ au vu des observations $Y_{0:n} = y_{0:n}$.

Ainsi :

$$\sum_{k=0}^{n-1} \xi_{x,x'}^k(y_{0:n}), \quad \sum_{k=0}^{n-1} \gamma_x^k(y_{0:n})$$

sont des estimateurs respectivement du nombre de transitions $x \rightarrow x'$ et du nombre de transitions issues de x dans la suite $X_{0:n}$, fondée sur l'observation de $Y_{0:n} = y_{0:n}$.

On peut alors proposer, partant du modèle $M = (\rho, Q, \psi)$, les "ré-estimateurs" suivants

$$\bar{\rho}(x) = \gamma_x^0(y_{0:n}), \quad (5.12)$$

$$\bar{Q}(x, x') = \frac{\sum_{k=0}^{n-1} \xi_{x,x'}^k(y_{0:n})}{\sum_{k=0}^{n-1} \gamma_x^k(y_{0:n})}, \quad (5.13)$$

$$\bar{\psi}(x, y) = \frac{\sum_{k=0}^n \mathbf{1}_{y_n=y} \gamma_x^k(y_{0:n})}{\sum_{k=0}^n \gamma_x^k(y_{0:n})}. \quad (5.14)$$

En effet il est clair que :

$$\begin{aligned} \bar{\rho}(x) &= \text{estimateur du nombre d'état } x \text{ à l'instant initial } k = 0, \\ \bar{Q}(x, x') &= \frac{\text{estimateur de nombre de transitions } x \rightarrow x'}{\text{estimateur du nombre de transitions issues de } x}, \\ \bar{\psi}(x, y) &= \frac{\text{estimateur de nombre de transitions issues de } x \text{ et émettant } y}{\text{estimateur du nombre de transitions issues de } x}. \end{aligned}$$

Ainsi, étant donné le modèle courant $M = (\rho, Q, \psi)$, l'utilisation des formules (5.12), (5.13) et (5.14) permet de ré-estimer le modèle et obtenir $\bar{M} = (\bar{\rho}, \bar{Q}, \bar{\psi})$. Baum a montré que :

- soit le modèle initial M est un point critique de la fonction de vraisemblance :

$$\nabla_M \mathcal{L}_n(M) = \nabla_M \mathbb{P}_M(Y_{0:n} = y_{0:n}) = 0$$

et alors $\bar{M} = M$;

- soit le modèle \bar{M} est strictement plus vraisemblable au sens où :

$$\mathcal{L}_n(\bar{M}) = P_{\bar{M}}(Y_{0:n} = y_{0:n}) > \mathcal{L}_n(M) = P_M(Y_{0:n} = y_{0:n}).$$

Itérer cette procédure donne un algorithme EM qui permet de déterminer un maximum local de la fonction de vraisemblance.

5.3 Algorithme de Viterbi et problème de reconnaissance

Etant donnés les paramètres $M = (\rho, Q, \psi)$ du modèle et une séquence d'observations $Y_{0:n}$, on désigne par le problème de reconnaissance la détermination de la séquence d'état $X_{0:n}$ qui est le plus vraisemblablement à l'origine de ces observations $Y_{0:n}$.

Quand il s'agit de calculer la loi conditionnelle de l'état à un instant donnée, les algorithmes "forward" et "backward" de Baum sont satisfaisants. En effet,

$$\mathbb{P}(X_k = x | Y_{0:n}) = \frac{q_k(x)}{\sum_{x' \in \mathcal{X}} q_k(x')}, \quad k \leq n,$$

pour tout $x \in \mathcal{X}$, où $q_k(x) = p_k(x)v_k(x)$.

En revanche, lorsqu'il s'agit de calculer des statistiques comme la moyenne conditionnelle, l'utilisation des lois conditionnelles précédentes s'avère sans sens vu que les états $\{1, \dots, N\}$ de la chaîne X sont des étiquettes et on ne peut pas effectuer des opérations algébriques sur ces valeurs. C'est ainsi que les statisticiens ont pensé à calculer des estimateurs comme *l'estimateur local du maximum a posteriori* tout en supposant que le maximum est atteint en un point unique :

$$X_k^{LMAP} = \arg \max_{x \in \mathcal{X}} \mathbb{P}(X_k = x | Y_{0:n}) = \arg \max_{x \in \mathcal{X}} q_k(x).$$

Cependant, un autre obstacle apparait. En effet, il peut arriver que la séquence $X_{0:n}^{LMAP}$ générée ne soit pas cohérente avec le modèle. Autrement dit, il peut arriver que $X_k^{LMAP} = x$, $X_{k+1}^{LMAP} = x'$ alors que la probabilité de transition $Q(x, x') = 0$. Pour cette raison, on opte finalement pour *l'estimateur global du maximum a posteriori*, défini par :

$$X_{0:n}^{MAP} = \arg \max_{x \in \mathcal{X}} \mathbb{P}(X_{0:n} = x_{0:n} | Y_{0:n}). \quad (5.15)$$

Cet estimateur est calculé à l'aide d'un algorithme de programmation dynamique, appelé *algorithme de Viterbi*.

Tout d'abord, remarquons qu'il est équivalent dans (5.15) de chercher l'argument du maximum de la loi conditionnelle ou de la loi jointe vu que la maximisation se fait sur $x_{0:n}$. On introduit alors la quantité :

$$\delta_x^n(y_{0:n}) = \max_{x_{0:n-1}} \mathbb{P}(X_{0:n-1} = x_{0:n-1}, X_n = x, Y_{0:n} = y_{0:n})$$

pour tout $n \geq 0$, $x_0, \dots, x_{n-1}, x \in \mathcal{X}$, $y_{0:n} \in \mathcal{Y}^{n+1}$. On définit, par la suite, la fonction valeur :

$$w_n(x) = \delta_x^n(Y_{0:n}), \quad n \geq 0, \quad x \in \mathcal{X}.$$

Théorème 5.3 (programmation dynamique). *La suite $(w_n)_{n \in \mathbb{N}}$ des fonctions valeur est solution de l'équation non-linéaire récurrente suivante :*

$$w_{n+1}(x) \leftarrow \psi(x, Y_{n+1}) \max_{x' \in \mathcal{X}} [Q(x', x) w_n(x')], \quad n \geq 0, \quad x \in \mathcal{X}, \quad (5.16)$$

$$w_0(x) \leftarrow \rho(x) \psi(x, Y_0), \quad x \in \mathcal{X}. \quad (5.17)$$

Preuve. D'après la proposition 5.1 :

$$\begin{aligned}\mathbb{P}(X_{0:n+1} = x_{0:n+1}, Y_{0:n+1} = y_{0:n+1}) \\ &= \rho(x_0)Q(x_0, x_1) \dots Q(x_n, x_{n+1})\psi(x_0, y_0) \dots \psi(x_{n+1}, y_{n+1}) \\ &= \psi(x_{n+1}, y_{n+1})Q(x_n, x_{n+1})\mathbb{P}(X_{0:n} = x_{0:n}, Y_{0:n} = y_{0:n})\end{aligned}$$

pour tout $n \geq 0$, $x_{0:n+1} \in \mathcal{X}^{n+2}$, $y_{0:n+1} \in \mathcal{Y}^{n+2}$. On en déduit que :

$$\max_{x_{0:n-1} \in \mathcal{X}^n} \mathbb{P}(X_{0:n+1} = x_{0:n+1}, Y_{0:n+1} = y_{0:n+1}) = \psi(x_{n+1}, y_{n+1})Q(x_n, x_{n+1})\delta_{x_n}^n(y_{0:n}).$$

La maximisation de cette dernière expression par rapport à x_n , nous donne :

$$\begin{aligned}\delta_{x_{n+1}}^{n+1}(y_{0:n+1}) &= \max_{x_{0:n} \in \mathcal{X}^{n+1}} \mathbb{P}(X_{0:n+1} = x_{0:n+1}, Y_{0:n+1} = y_{0:n+1}) \\ &= \max_{x_n \in \mathcal{X}} \max_{x_{0:n-1} \in \mathcal{X}^n} \mathbb{P}(X_{0:n+1} = x_{0:n+1}, Y_{0:n+1} = y_{0:n+1}) \\ &= \psi(x_{n+1}, y_{n+1}) \max_{x_n \in \mathcal{X}} [Q(x_n, x_{n+1})\delta_{x_n}^n(y_{0:n})]\end{aligned}$$

d'où le résultat. \square

Corollaire 5.1 (Algorithme de Viterbi). *Soit n un instant fixé, on commence par calculer la séquence $w_{0:n}$. On obtient, par la suite, l'estimateur $X_{0:n}^{MAP}$ par récurrence rétrograde :*

$$X_n^{MAP} \leftarrow \arg \max_{x \in \mathcal{X}} w_n(x) \quad (5.18)$$

puis, pour $k = n - 1, \dots, 0$:

$$X_k^{MAP} \leftarrow \arg \max_{x \in \mathcal{X}} [Q(x, X_{k+1}^{MAP}) w_k(x)]. \quad (5.19)$$

Preuve. Par définition, $w_k(x)$ représente la probabilité maximale d'avoir l'état x à l'instant k en maximisant par rapport à toutes les trajectoires possibles $X_{0:k-1}$. Ainsi, parmi toutes les trajectoires qui aboutissent dans l'état x' à l'instant $k + 1$, la trajectoire de plus grande probabilité est passée dans l'état :

$$I_k(x') = \arg \max_{x \in \mathcal{X}} [Q(x, x')w_k(x)]$$

à l'instant k précédent, tout en supposant comme toujours que ce maximum est atteint en un point unique. En outre, on a nécessairement :

$$Q(I_k(x'), x') > 0$$

ce qui garantit que la transition de l'état $I_k(x')$ vers l'état x' est possible pour le modèle. La trajectoire optimale est calculée de la façon suivante : on remarque d'abord que

$$\begin{aligned}\max_{x_{0:n} \in \mathcal{X}^{n+1}} \mathbb{P}(X_{0:n} = x_{0:n}, Y_{0:n} = y_{0:n}) &= \max_{x_n \in \mathcal{X}} \max_{x_{0:n-1} \in \mathcal{X}^n} \mathbb{P}(X_{0:n} = x_{0:n}, Y_{0:n} = y_{0:n}) \\ &= \max_{x_n \in \mathcal{X}} \delta_{x_n}^n(y_{0:n}).\end{aligned}$$

On en déduit (5.18) et d'une manière récurrente (5.19). \square

En conclusion, ce chapitre présente le modèle du Markov caché et les problèmes canoniques liés. En partant d'une séquence d'observations, une première étape consiste à résoudre le problème de ré-estimation et à déterminer les paramètres du modèle qui correspond mieux à nos observations. Ensuite, la deuxième étape est la résolution du problème de reconnaissance et la détermination par conséquent de la séquence d'états cachés. Ces deux problèmes se résolvent respectivement par l'algorithme de Baum-Welsh et l'algorithme de Viterbi. Le chapitre suivant présente l'application de ce modèle sur nos données. En considérant les log-rendements comme la séquence d'observations, on détermine les paramètres du modèle dans un premier temps et la séquence d'états à valeur crise ou non-crise dans un deuxième temps.

Chapitre 6

Détection des régimes : régime crise / régime hors crise

Depuis son apparition, le MMC a eu des divers applications en finance. En effet, la variable cachée représente une information cachée dans un signal donné. Elle est intéressante dans la mesure où elle fournit un résumé de l'information plus simple et plus fondamental que les données elles-mêmes. Ainsi la chaîne d'un MMC peut avoir une dynamique plus simple que le signal. Par exemple, l'introduction d'une chaîne de Markov cachée dans les modèles A.R.C.H a permis de capter les soudains retournements du niveau de volatilité [17].

Comme dans toute approche de classification non supervisée, la question essentielle est de savoir quelle genre d'information on souhaite extraire des données. Ainsi, parmi les études récentes appliquant ce concept, on trouve celle de Y. Zhang [28] où la prédiction des séries temporelles des rendements d'un portefeuille était l'objet de l'application du MMC. Nous citons aussi la thèse de N. Mayo [21] où il s'est intéressé à la détection d'arbitrage et la prévision des volumes. Dans ses travaux, il a utilisé les modèles à variables cachés (MVC), la généralisation des MMC. L'hypothèse de markovianité dans le MMC a pour unique justification le caractère calculatoire. Il est en effet possible de calculer la vraisemblance en un temps linéaire par rapport au nombre d'observations, alors qu'il est à priori exponentiel. Ce gain en temps de calcul est souvent vital en pratique.

Dans notre cas, l'objectif est toujours de proposer un indicateur de risque optimisé en termes de performance et de précision. Ainsi, nous choisissons, comme information à extraire, les changements de régime et nous ajoutons, par conséquent, un nouveau critère de détection de crise dans notre calcul de la Value-at-Risk. Pour cela, on suppose l'existence de deux régimes sur le marché financier : un régime qui décrit les périodes de crise et un deuxième pour les périodes normales, dite dans ce rapport périodes de non-crise. On cherche à classer les données et à attribuer à chaque jour une étiquette crise ou non-crise par l'application du modèle de Markov caché.

6.1 Classification de données boursières

La classification a été appliquée sur les log-rendements journaliers du cours boursier observé sur le marché. Une première étape a utilisé l'algorithme k -means, un algorithme basique qui cherche à diviser les données en classes tout en minimisant la distance intra-classe et en maximisant la distance inter-classes. Le résultat n'est pas satisfaisant et on voit visuellement dans la figure 6.1 l'absence de la notion de période. C'est ainsi qu'on a eu recours à une modélisation plus sophistiquée et on a opté pour le modèle de Markov caché.

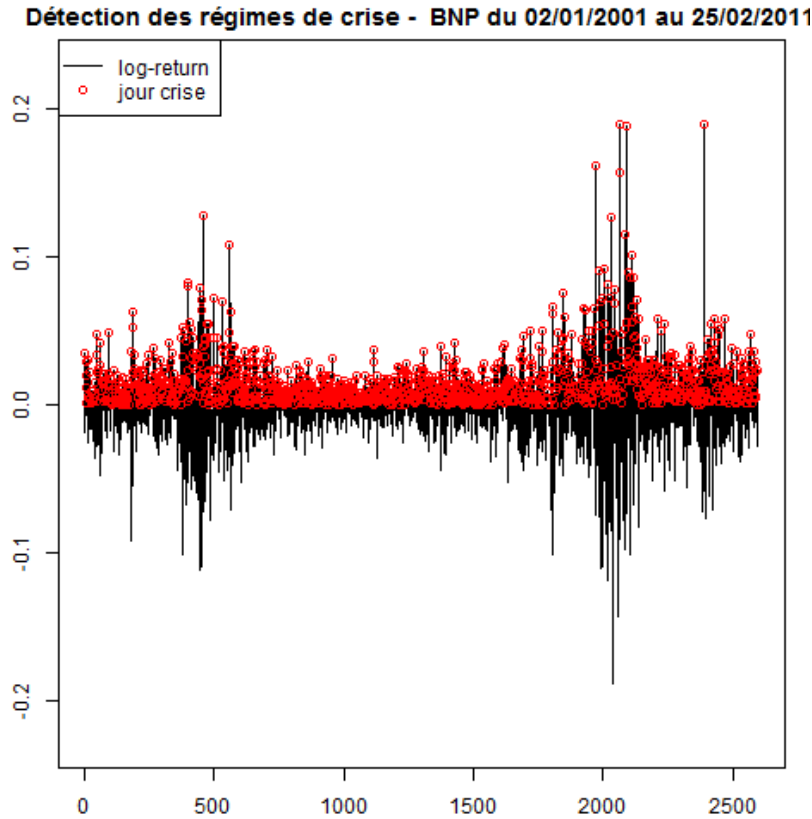


FIGURE 6.1 – Classification et détermination des périodes crise et des périodes non-crise pour l'actif BNP en utilisant l'algorithme k -means.

On considère la série des log-rendements $Y_{1:n} = (Y_i, 1 \leq i \leq n)$ comme étant la séquence d'observations dans notre modèle. On suppose, par la suite, l'existence d'une chaîne de Markov cachée $X_{1:n} = (X_i, 1 \leq i \leq n)$ à valeurs dans $\mathcal{X} = \{c, n\}$. Cette chaîne représente la séquence des états de notre modèle et elle indique si l'on est dans un jour crise (c) ou dans un jour normal (n).

Notre travail se décompose alors en trois parties : Tout d'abord, on cherche à estimer les paramètres du modèle $M = (\rho, Q, \psi)$ qui maximise la vraisemblance d'apparition de $Y_{1:n}$

(problème de ré-estimation). Ensuite, étant donnés les paramètres du modèle M et les observations $Y_{1:n}$, on détermine la série des états $X_{1:n}$ qui est le plus vraisemblablement à l'origine de $Y_{1:n}$ (problème de reconnaissance). Finalement, on propose une nouvelle méthode de calcul de la Value-at-Risk qui tient compte de cette classification.

6.2 Problème de ré-estimation

On dispose d'une série de log-rendements $Y_{1:n}$. On applique l'algorithme de Baum-Welsh pour déterminer les paramètres $M = (\rho, Q, \psi)$ du modèle de Markov caché. Les résultats sont les suivants pour le cas de l'actif BNP :

- La loi initiale ρ de X :

$$\rho(c) = 4,97 \cdot 10^{-05}, \quad \rho(n) = 0.9999503.$$

Il est tout à fait logique d'avoir $\rho(n)$ presque égale à 1. En effet, $\rho(n) = \mathbb{P}(X_1 = n)$ est la probabilité que le premier jour de la séquence soit en période normale (hors-crise).

- La matrice de transition Q de X :

$$\begin{pmatrix} Q(c, c) & Q(c, n) \\ Q(n, c) & Q(n, n) \end{pmatrix} = \begin{pmatrix} 0.9795115 & 0.0204885 \\ 0.0069161 & 0.9930839 \end{pmatrix}.$$

On voit bien que la probabilité de passer d'une période de crise à une période normale est 3 fois plus élevée que celle de passer d'une période normale à une période crise. En effet, la première quantité s'interprète par la probabilité $Q(c, n)$ d'être aujourd'hui en période normale sachant que la veille on était en crise, alors que la deuxième quantité est définie par la probabilité $Q(n, c)$ d'être aujourd'hui en crise sachant que la veille on était en période normale. De l'autre côté, la probabilité de rester en état hors crise est supérieure à celle de rester en état crise. Ces résultats, en terme de matrice de transition de X , sont acceptables et conformes avec la réalité du marché financier.

- La moyenne et la variance par ligne de la matrice d'émission ψ sont :

$$\begin{pmatrix} moy(\psi(c, Y)) & var(\psi(c, Y)) \\ moy(\psi(n, Y)) & var(\psi(n, Y)) \end{pmatrix} = \begin{pmatrix} -0.0012500 & 0.0018043 \\ 0.0005150 & 0.0002108 \end{pmatrix}.$$

La variance de la série des log-rendements ayant l'état crise est beaucoup plus élevée que celle des log-rendements ayant l'état hors-crise (normal). Cela s'explique par la grande variance qui caractérise les périodes de crise dans les marchés financiers.

En terme de performance informatique, le nombre moyen d'itérations, effectuées par l'algorithme, pour aboutir au résultat (les paramètres du modèle qui présentent un maximum local en terme de mieux correspondre aux log-rendements) est 15 itérations. Cette moyenne est calculée en lançant l'algorithme 10 fois sur les log-rendements de l'actif BNP.

En conclusion, les paramètres de notre modèle n'ont pas de contradiction avec la réalité du fonctionnement du marché financier. En plus, des résultats similaires ont été trouvés lors de l'application du modèle de Markov caché sur d'autres actifs boursiers. On peut dire alors que la résolution du problème de ré-estimation a donné des résultats conformes avec ce qu'on attend. L'étape suivante est la résolution du problème de reconnaissance et la détermination de la série des états, qui fait l'objet de la prochaine section.

6.3 Problème de reconnaissance

Ayant maintenant une série d'observation $Y_{1:n}$ représentant les log-rendements d'un actif boursier et les paramètres $M = (\rho, Q, \psi)$ calculés par l'algorithme de Baum-Welsh, on cherche à déterminer la série des états $X_{1:n}$ et à attribuer ainsi à chaque jour une étiquette crise (c) ou non-crise (n). Ce problème est connu sous le nom du problème de reconnaissance et il est résolu par l'algorithme de Viterbi.

La figure 6.2 montre le résultat de l'application de cet algorithme sur les log-rendements du BNP. On voit en rouge les périodes de crise et en noir les périodes normales, dite de non-crise.

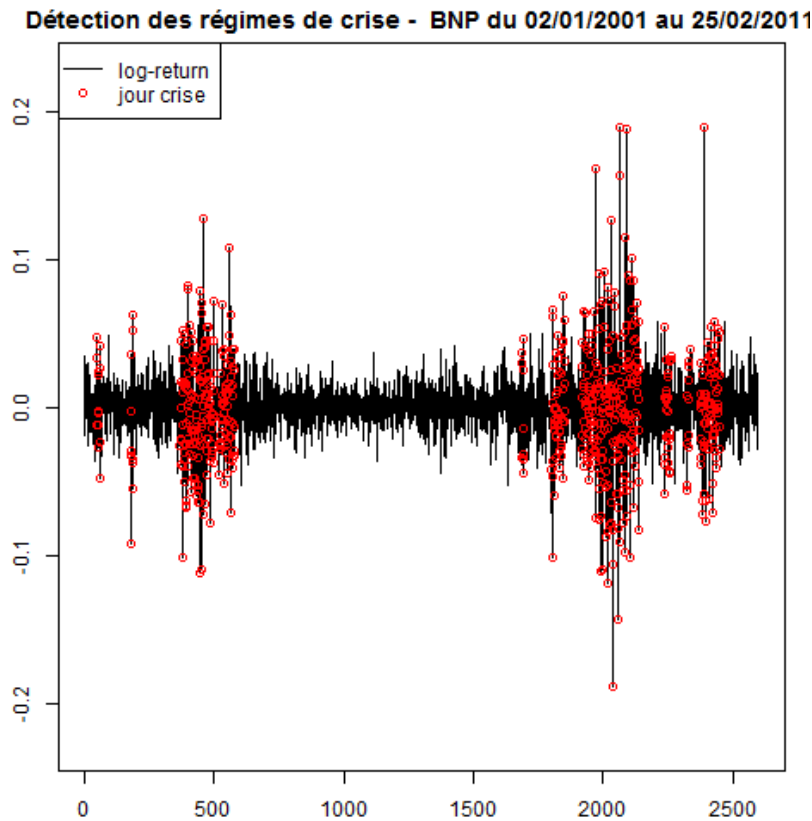


FIGURE 6.2 – Classification et détermination des périodes crise et des périodes non-crise pour l'actif BNP.

6.4 Etude des crises

L'un des apports de notre classification est le fait de pouvoir dire si aujourd'hui on est en crise ou en période normale. Les deux principales informations qu'on peut dégager sont la durée moyenne d'une crise et la distribution du temps de retour à la crise. En se basant sur ces informations, on cherche à trouver des classes de comportement d'actifs boursier.

En étant dans une période de non-crise, le modèle le plus intuitive pour le temps de retour en crise τ est celui d'une loi géométrique de probabilité de succès

$$p = \mathbb{P}(X_{i+1} = c / X_i = n) = Q(n, c).$$

Ainsi, la probabilité que le retour en crise soit dans k jours est :

$$\mathbb{P}(\tau = k) = (1 - p)^{k-1} p = Q(n, n)^{k-1} Q(n, c)$$

La densité théorique de la distribution de temps de retour en crise pour l'actif BNP Paribas est donnée par la figure 6.3. Cependant, numériquement on ne dispose que de 8 temps de retour en crise vu que les données ne couvrent qu'une période de 10 ans.

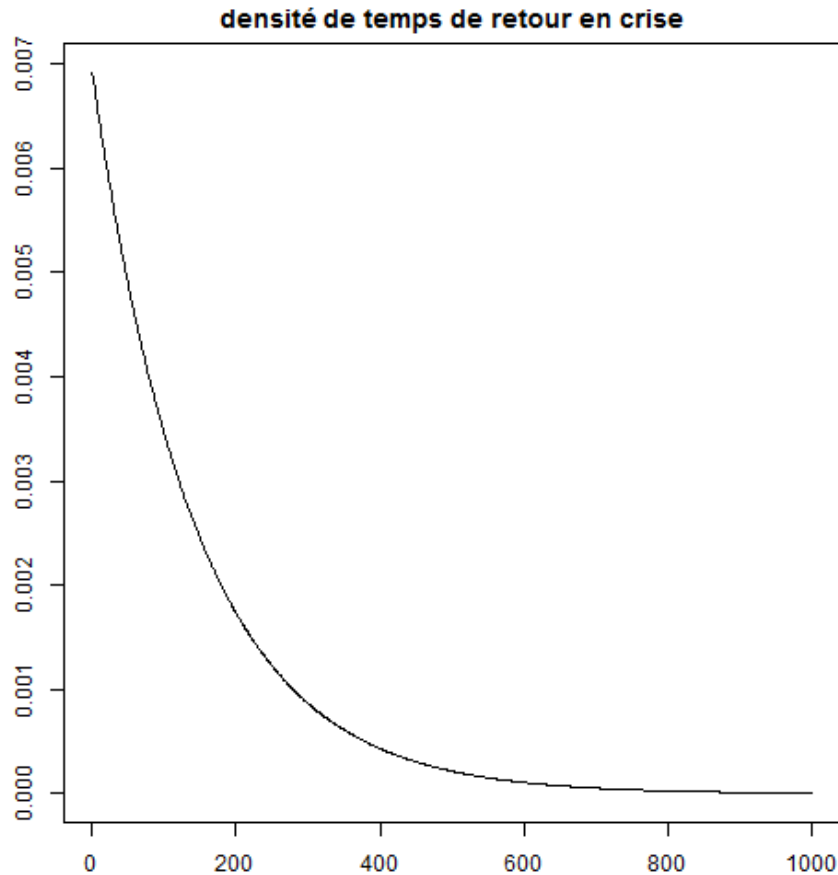


FIGURE 6.3 – Densité de temps de retour en crise pour l'actif BNP

6.5 Calcul de la VaR

L'indicateur VaR est l'objet central de cette étude. Nous proposons dans cette section une nouvelle méthode de calcul de la VaR. Cette méthode repose sur l'idée de profiter de l'information crise/non-crise obtenue suite à la classification. D'habitude, on estime la VaR à l'instant t par les 252 jours qui précèdent. Notre méthode repose sur les 252 jours précédents ayant le même état. Les étapes sont les suivantes :

- On cherche l'instant t_0 tel qu'il existe au moins 252 jours d'état crise et 252 jours d'état non-crise précédant cet instant. Au delà de cet instant, on est sûr d'avoir suffisamment de données pour l'estimation.
- Pour $t \geq t_0$, on calcule la VaR à l'instant t de la façon suivante : On suppose que l'état de l'instant t est celui de la veille (instant $t - 1$), une supposition qui reste raisonnable vu que les probabilités de changement de régime sont presque nulles. Ensuite, on estime la VaR à l'aide de 252 jours passés ayant le même état que le jour t .

L'avantage de cette méthode réside dans l'estimation avec des données semblables. En effet, la VaR est définie comme la perte maximale à subir en dehors d'un événement de probabilité p en un horizon temporel t . Il s'avère pertinent de calculer cette perte maximale avec des données de même état, ayant par conséquent le même comportement. Ainsi, l'estimation des paramètres de la loi puissance, et de la VaR par la suite, se fait par les 252 données précédentes de même régime.

Les figures 6.4 et 6.5 montrent les résultats du backtesting qui nous permettent de comparer la méthode standard et la nouvelle méthode basée sur l'utilisation de la classification. On remarque une amélioration du backtesting pour certains actifs. A titre d'exemple, pour les actifs CNP et GLE, le backtesting avec la méthode de classification, donne des résultats très proches de 1% avec des intervalles de confiance acceptables. Des résultats sont aussi prouvés pour les actifs CA et TEC. D'autre part, pour certains actifs comme BNP, FGR, LI et RI, le backtesting ne donne pas le résultat attendu, avec des intervalles de confiance supérieurs à 1%. Cela est dû au fait que pour ces derniers actifs, la VaR a été calculée avec un nombre limité de mesures. En effet, notre méthode de calcul de VaR repose sur le fait qu'il existe un instant t_0 à partir duquel on commence notre estimation. On a aussi remarqué que pour ces derniers actifs cet instant se trouve dans le dernier tier des données journalières, d'où un nombre de données très limité. Ce qui explique le backtesting non-concluant pour ces cas.

Autrement, un effet de recul dans le passé est aussi observable dans les figures 6.4 et 6.5. En effet, la VaR calculée par la méthode de classification, pour tous les actifs, diffère de la VaR calculée par la méthode standard par l'apparition de nombreuses pics tout au long de la courbe du seuil de 1%. L'interprétation qu'on donne, est qu'au début d'une période crise ou une période normale, on n'a pas suffisamment de données ayant le même état qui précèdent directement l'instant de calcul. Ainsi, on recule dans le passé jusqu'à

rassembler 252 données de même état. Cela peut provoquer parfois des pics dans la VaR lorsque les données utilisées pour la calculer ont une variance plus importante. Ces pics représentent des sur-évaluations ponctuelles de la VaR à certains instants. Donc, elles ne peuvent pas être considérées comme un inconvénient. Il s'agit juste de limiter la liberté des clients de risquer sur l'actif pour les jours en question.

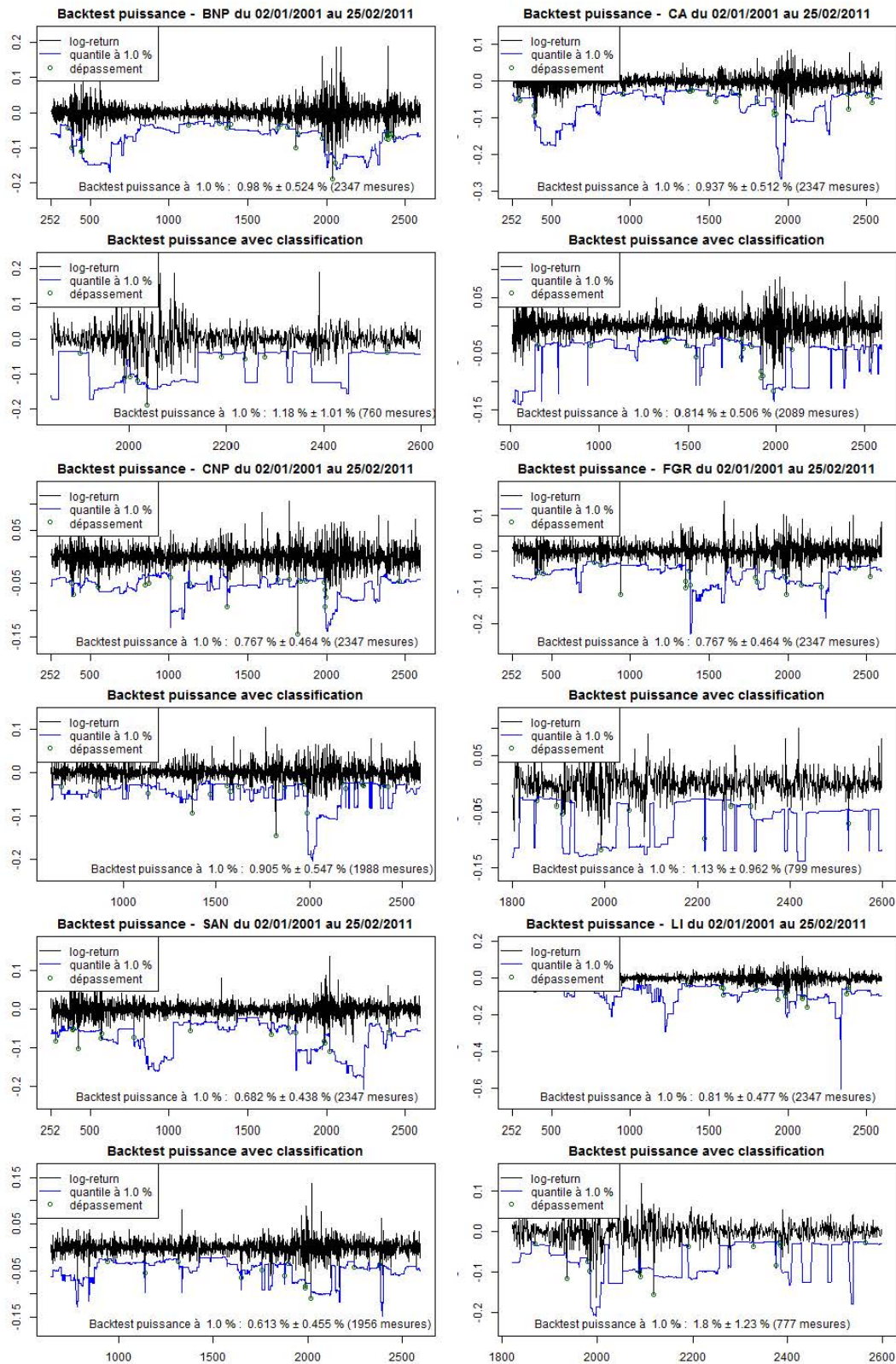


FIGURE 6.4 – Comparaison du backtesting à 1% sur la VaR calculée par la méthode standard et par la méthode de calssification sur les actifs BNP, CA, CNP, FGR, SAN, LI.

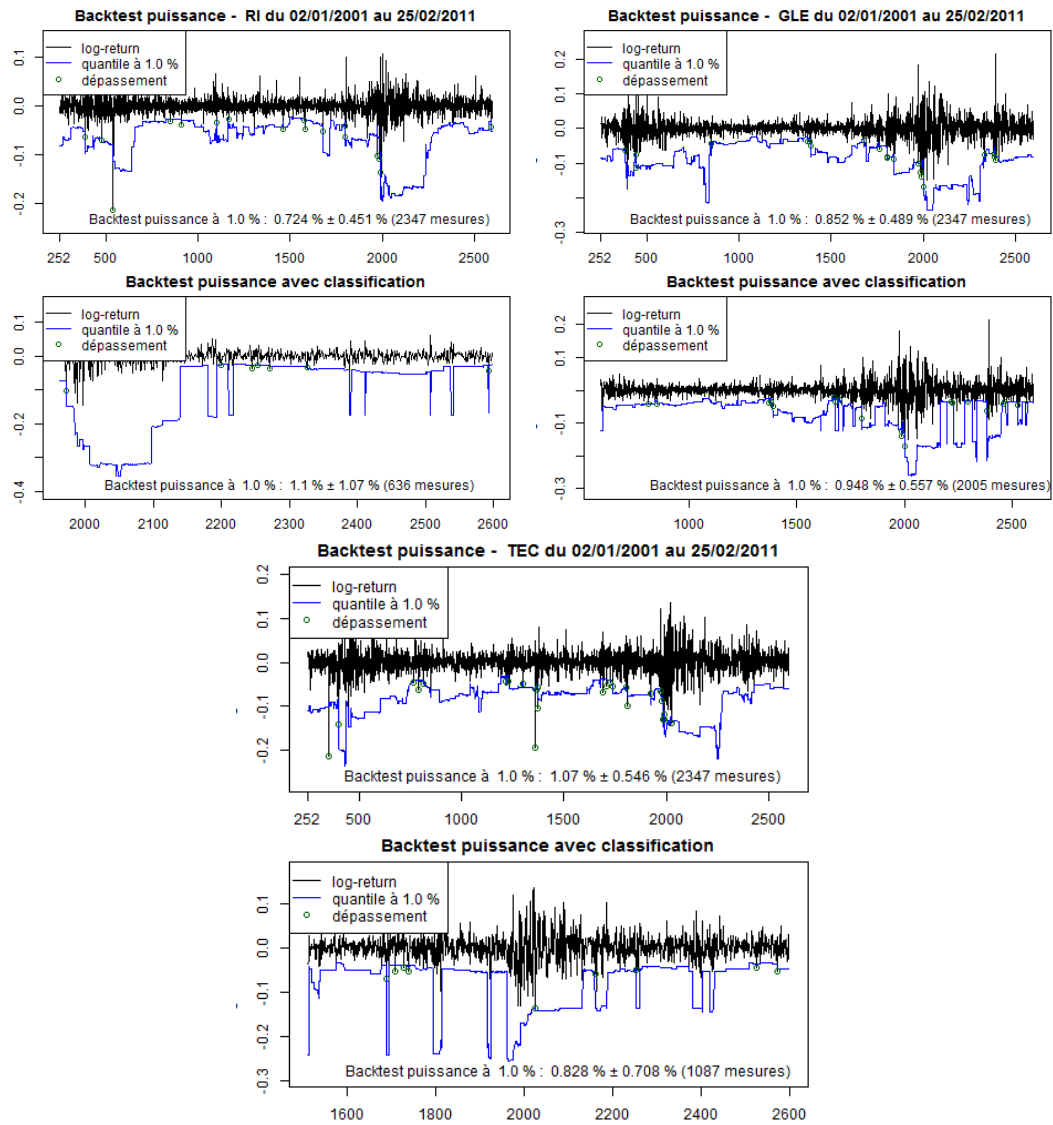


FIGURE 6.5 – Comparaison du backtesting à 1% sur la VaR calculée par la méthode standard et par la méthode de calssification sur les actifs RI, GLE, TEC.

Chapitre 7

Conclusion

Ce rapport s'inscrit dans le cadre de la gestion de risque. Il s'agit de proposer un indicateur de risque capable de prédire les pertes sur les actifs financiers. Nous étudions le marché financier sous un modèle des loi puissances et nous introduisons la VaR, notre indicateur de risque. Nous définissons, par la suite, notre approche de calcul de la VaR basée sur une classification des données selon le critère de crise.

La première partie de ce manuscrit présente le modèle des lois puissances adopté dans cette étude pour décrire les pertes logarithmiques. On présente les estimateurs des paramètres de la loi puissance, ainsi que la VaR sous ce modèle. Une analyse de la sensibilité de ces estimateurs à certains facteurs extérieurs est expliquée ensuite. La deuxième partie s'intéresse au modèle de Markov caché, qui permet d'associer à une série d'observations une série d'état cachés, et à la classification des données selon le critère de crise. Cette classification nous a permis de proposer une nouvelle méthode de calcul de la VaR : On estime la VaR à l'instant t en utilisant les données des 252 jours précédents ayant le même état que l'instant t .

L'application de cette méthode de calcul sur les données boursières améliore les résultats du test de validation de la VaR pour plusieurs actifs financier. Par ailleurs, pour d'autres actifs, les résultats ne sont pas concluantes. En effet, pour intégrer la classification dans le calcul de la VaR, on est obligé de commencer à partir d'un certain rang de données, et pour quelques actifs ce rang est tellement grand que le nombre de données utilisées ne permettent pas d'obtenir un résultat performant. En conclusion, notre approche de calcul de la VaR surmonte le nombre limité de données, dans plusieurs cas, et engendre des améliorations dans le backtesting.

Les perspectives directes de ce travail sont d'explorer mieux les caractéristiques de la crise, à savoir la durée moyenne de crise et la distribution de temps de retour. On cherchera, par exemple, à tester si les temps de retour en crise suivent réellement la distribution géométrique mentionnée. Le but est de dégager des classes de comportement d'actifs selon ces critères afin de mieux comprendre le fonctionnement du marché

financier. En perspectives indirectes, nous trouvons toujours le problème multidimensionnel et l'application de nos résultats sur un portefeuille composé de plusieurs actifs financiers. Ce problème peut être formulé de la façon suivante : déterminer une fonction qui donne à chaque combinaison linéaire d'actifs financiers une Value-at-Risk associée. Cela revient à étudier les combinaisons des lois puissances et à construire une forme générale ou asymptotique de la distribution en question.

Bibliographie

- [1] Leonard E. BAUM : An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *In Inequalities, III (Proc. Third Sympos., Univ. California, Los Angeles, Calif., 1969; dedicated to the memory of Theodore S. Motzkin)*, pages 1–8. Academic Press, New York, 1972.
- [2] Leonard E. BAUM et J. A. EAGON : An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bull. Amer. Math. Soc.*, 73:360–363, 1967.
- [3] Leonard E. BAUM et Ted PETRIE : Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, 37:1554–1563, 1966.
- [4] Leonard E. BAUM, Ted PETRIE, George SOULES et Norman WEISS : A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.*, 41:164–171, 1970.
- [5] Leonard E. BAUM et George R. SELL : Growth transformations for functions on manifolds. *Pacific J. Math.*, 27:211–227, 1968.
- [6] Jan BEIRLANT, Yuri GOEGEBEUR, Jozef TEUGELS et Johan SEGERS : *Statistics of extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2004. Theory and applications, Avec la contribution de Daniel De Waal et Chris Ferro.
- [7] Jeremy BERKOWITZ : Testing density forecasts, with applications to risk management. *J. Bus. Econom. Statist.*, 19(4):465–474, 2001.
- [8] Souhail BOUKHEROUAA, Nicolas CHAMPAGNAT, Madalina DEACONU et Antoine LEJAY : Mesure de risques : calcul de la Value-at-Risk et application à la gestion de portefeuilles. Contrat, janvier 2013.
- [9] S.D. CAMPBELL : *A review of backtesting and backtesting procedures*. Finance and economics discussion series. Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, 2005.
- [10] C. CHATFIELD : Calculating interval forecasts. *Journal of Business and Economics statistics*, 11:121–135, 1993.
- [11] P. CHRISTOFFERSEN et D. PELLETIER : Backtesting Value-at-Risk : A Duration-Based Approach. *Journal of Financial Econometrics*, 2:84–108, 2004.

- [12] Peter F. CHRISTOFFERSEN : Evaluating interval forecasts. *Internat. Econom. Rev.*, 39(4):841–862, 1998. Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance.
- [13] John COTTER et Yan Ping ZHONG : *Validating backtests of risk measures*. Centre for Financial Markets, University College Dublin, 2007.
- [14] A. P. DEMPSTER, N. M. LAIRD et D. B. RUBIN : Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–38, 1977. With discussion.
- [15] Robert J. ELLIOTT, Lakhdar AGGOUN et John B. MOORE : *Hidden Markov models*, volume 29 de *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1995. Estimation and control.
- [16] E. HAEUSLER et J. L. TEUGELS : On asymptotic normality of Hill’s estimator for the exponent of regular variation. *Ann. Statist.*, 13(2):743–756, 1985.
- [17] James D. HAMILTON : Autoregressive conditional heteroskedasticity and changes in regime. *J. Econometrics*, 64(1):307–333, 1994.
- [18] Bruce M. HILL : A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3(5):1163–1174, 1975.
- [19] P. H. KUPIEC : Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3:73–84, 1995.
- [20] Iain L. MACDONALD et Walter ZUCCHINI : *Hidden Markov and other models for discrete-valued time series*, volume 70 de *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1997.
- [21] N. MAYO : *Les modèles à variables cachées et leurs applications en finance : risque systématique, détection d’arbitrage et prévision des volumes*. Thèse de doctorat, Université Paris I Panthéon-Sorbonne, 2010.
- [22] Geoffrey J. MCLACHLAN et Thriyambakam KRISHNAN : *The EM algorithm and extensions*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second édition, 2008.
- [23] Alfréd RÉNYI : On the theory of order statistics. *Acta Math. Acad. Sci. Hungar.*, 4:191–231, 1953.
- [24] Sidney I. RESNICK : *Extreme values, regular variation, and point processes*, volume 4 de *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York, 1987.
- [25] Khaled SALHI : Modélisation et estimation des valeurs extrêmes sur les marchés financiers. Projet d’initiation à la recherche master 2 MFA, Nancy, février 2013.
- [26] Gennady SAMORODNITSKY et Murad S. TAQQU : *Stable non-Gaussian random processes*. Stochastic Modeling. Chapman & Hall, New York, 1994. Stochastic models with infinite variance.
- [27] Ishay WEISSMAN : Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.*, 73(364):812–815, 1978.

- [28] Y. ZHANG : *Prediction of financial time series with hidden Markov models*. Thèse de doctorat, Shandong University, China, 2001.