



HAL
open science

Étude spatio-temporelle de la dynamique des rendements boursiers au travers de la théorie d'information et de la mécanique statistique

Geoffrey Ducournau-Voisin

► **To cite this version:**

Geoffrey Ducournau-Voisin. Étude spatio-temporelle de la dynamique des rendements boursiers au travers de la théorie d'information et de la mécanique statistique. Economies et finances. Université Montpellier, 2021. Français. NNT : 2021MONTD033 . tel-03692859

HAL Id: tel-03692859

<https://theses.hal.science/tel-03692859>

Submitted on 10 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Sciences Économiques

École doctorale Économie et Gestion

Unité de recherche MRE

**Étude spatio-temporelle de la dynamique des rendements boursiers
au travers de la théorie d'Information et de la Mécanique Statistique**

Présentée par Geoffrey DUCOURNAU – VOISIN

Le 13 décembre 2021

Sous la direction de Mme Françoise SEYTE

Devant le jury composé de

Mme Françoise SEYTE, Maître de Conférences HDR à l'Université de Montpellier

M. Stéphane MUSSARD, Professeur à l'Université de Nîmes

M. Matthieu GARCIN, Professeur à l'École Supérieure d'Ingénieurs Léonard de Vinci

M. Roman MESTRE, Maître de Conférences à l'Université de Montpellier

M. Walter BRIEC, Professeur à l'Université de Perpignan

M. Zied FTITI, Professeur à EDC Paris Business School

M. Daniel MELHEM, Directeur de la société de recherche Dimtech S.A.S

Directrice de thèse

Président du jury

Examineur

Examineur

Rapporteur

Rapporteur

Invité



UNIVERSITÉ
DE MONTPELLIER

L'université n'entend donner aucune approbation ni improbation aux opinions émises dans la thèse ; ces opinions doivent être considérées comme propres à leurs auteurs

Dédicace

Je dédicace cette thèse à :

Ma tendre mère, mon porte chance, qui a façonné la personne que je suis, et à qui je serai à jamais redevable devant tant de sacrifices et de dévotions.

Mon cher grand père, une source de motivation, une source d'inspiration, un modèle qui n'a cessé et ne cessera de déterminer les prochaines étapes de ma vie.

Mon épouse, qui ne cesse de m'épauler, de me soutenir et de m'encourager dans la poursuite de mes choix.

A toute ma famille et à tous mes amis

Remerciements

Au terme de ma thèse, je tiens à exprimer mes sincères remerciements à ma directrice de thèse, Madame Françoise SEYTE, qui m'a permis d'entreprendre une voie académique tant désirée et qui a su me faire confiance en me laissant une grande flexibilité dans mon choix de poursuivre des travaux de recherches interdisciplinaires.

Je remercie également Monsieur Daniel MELHEM, directeur de la société Dimtech S.A.S grâce et avec qui, pendant ces années de thèse, j'ai pu évoluer à la fois sur le plan professionnel, humain et spirituel. Il a su m'épauler, me soutenir et m'encourager. Sans lui cette thèse n'aurait pas été possible.

Je remercie la société DIMTECH ainsi que l'ANRT qui ont rendu mon parcours de doctorant possible en finançant pendant ces trois années ma thèse et mes travaux de recherche sous la forme d'un contrat CIFRE.

J'adresse tous mes remerciements à Monsieur Stéphane MUSSARD, Professeur à l'Université de Nîmes qui a su à de nombreuses reprises me prodiguer des conseils dans l'avancement de mes recherches. Je le remercie également pour siéger comme directeur de jury de ma soutenance de thèse.

J'adresse mes remerciements à Monsieur Mathieu GARCIN, professeur à l'école supérieure d'ingénieurs Léonard de Vinci pour l'intérêt qu'il porte à mes travaux en acceptant de siéger parmi les membres du jury en qualité d'examineur.

J'adresse mes remerciements à Monsieur Walter BRIEC, professeur à l'Université de Perpignan, et Monsieur Zied FTITI, professeur à l'école EDC Paris Business School, de l'honneur qu'ils m'ont fait en acceptant d'être rapporteurs de cette thèse.

J'adresse mes remerciements aux membres du laboratoire Montpellier Recherche en Économie (MRE) pour l'accueil qu'ils m'ont accordé.

Enfin, je remercie tous ceux qui, de près ou de loin, ont contribué à l'aboutissement de cette thèse.

Résumé

Dans cette thèse, nous nous intéressons à l'étude et à la modélisation du phénomène de complexité émergeant des séries financières selon des échelles temporelles et spatiales différentes.

Nous proposons une méthode statistique pour mesurer le niveau de complexité d'une série financière quelconque sans aucun a priori sur le caractère déterministe ou aléatoire de cette dernière. L'étude spatio-temporelle repose sur une approche statistique connue sous le nom de la Dynamique Symbolique, une branche de la théorie d'information permettant notamment de simplifier l'analyse de l'évolution des systèmes dynamiques complexes au travers d'une transformation des données observées en une suite de symboles. Nous en proposons une application sur l'indice Standard & Poor500 et constatons, au travers de l'entropie généralisée de Shannon, que sa dynamique varie fortement selon l'échelle de temps d'étude, exhibant une information plus prévisible sur des fréquences temporelles plus hautes et se caractérisant par un processus plus Markovien sur des échelles plus basses.

Puis nous montrons par des expériences numériques que le phénomène complexe de cette série financière est la conséquence d'une superposition de plusieurs statistiques locales sur des échelles de temps différentes, se traduisant à l'échelle globale par l'émergence d'une variance non constante.

Nous proposons ainsi une méthode de modélisation de cette variance non constante par une généralisation de la statistique de Boltzmann au travers d'une inférence bayésienne. Nous montrons notamment que les familles inverses gamma et log normal caractéristiques d'une loi de puissance et de présence de mémoire longue permettent de modéliser de manière significative le processus de volatilité stochastique sur des échelles de hautes fréquences. Sur des échelles de basses fréquences, la famille gamma est préférée. Nous montrons par ailleurs que cette transition de statistique selon l'échelle temporelle permet entre autres d'expliquer le changement du comportement complexe de la série étudiée, comme la présence de forte corrélation entre les rendements observés sur des hautes fréquences et le caractère plus aléatoire et moins prévisible sur des échelles plus basses.

Mots clés : Système Complexe, Dynamique Symbolique, Théorie de l'Information, Mécanique Statistique, Inférence Bayésienne, Volatilité Stochastique.

Abstract

We are interested in the study and the modeling of the phenomenon of complexity emerging from financial series according to different time scales and spaces.

We propose a statistical method to measure the level of complexity of any financial timeseries without any a priori on the deterministic or random nature of the latter. The spatio-temporal study is based on a statistical approach known under the name of Dynamic Symbolic, a branch of information theory allowing to simplify the analysis of the evolution of complex dynamic systems through the transformation of the observed data in a series of symbols. We propose an application of it on the Standard & Poor500 index and note through the generalized entropy of Shannon that its dynamics vary strongly according to the time scale of study, showing more predictable information on higher temporal frequencies and characterized by a more Markovian process on lower scales.

We show by numerical experiments that the complex phenomenon of this financial timeseries is the consequence of a superposition of several local statistics on different time scales resulting on the global scale by the emergence of a non-constant variance.

We thus propose a method of modeling this non-constant variance by a generalization of the Boltzmann's statistics through a Bayesian inference. We show that the inverse gamma and log normal family's characteristic of a power-law and the presence of long memory allow us to significantly model the process of stochastic volatility on high frequency scales. On low frequency scales, the gamma family is preferred. We also show that this transition of statistics according to the temporal scale makes it possible, among other things, to explain the change in the complex behavior of the studied series, such as the presence of a strong correlation between the returns observed on high frequencies and the more random and less predictable characteristics on lower scales.

Keywords: Complex System, Dynamic Symbolic, Information Theory, Statistical Mechanics, Bayesian Inference, Stochastic Volatility.

Sommaire

INTRODUCTION	6
CHAPITRE 1 : SYSTÈME COMPLEXE	13
1.1. INTRODUCTION	13
1.2. PROPRIÉTÉS ET CARACTÉRISTIQUES D'UN SYSTÈME COMPLEXE.....	16
1.3. LE COMPORTEMENT FRACTAL DES SYSTÈMES COMPLEXES	27
1.4. DIFFÉRENCES ENTRE COMPLEXITÉ STATISTIQUE ET COMPLEXITÉ DÉTERMINISTE	31
1.5. CONCLUSION	36
CHAPITRE 2 : MESURE DE LA COMPLEXITÉ : UNE APPROCHE STATISTIQUE.	39
2.1 INTRODUCTION	39
2.2 LES DIMENSIONS FRACTALES.....	41
2.3 DYNAMIQUE SYMBOLIQUE ET THÉORIE DE L'INFORMATION DE SHANNON.....	47
2.4 DIMENSION ET ENTROPIE GÉNÉRALISÉE.....	55
2.5 ANALYSE ET MESURE DE LA COMPLEXITÉ D'UNE SÉRIE ÉCONOMIQUE AU TRAVERS DE LA DYNAMIQUE SYMBOLIQUE ET DE LA THÉORIE DE L'INFORMATION	66
2.6 CONCLUSION	93
CHAPITRE 3 : MODÉLISATION DE LA COMPLEXITÉ : APPROCHE SUPER STATISTIQUE	96
3.1 INTRODUCTION	96
3.2 SUPER-STATISTIQUES : PRINCIPES FONDATEURS	98
3.3 SUPERPOSITION DE STATISTIQUES DES SÉRIES FINANCIÈRES : EXPÉRIENCES NUMÉRIQUES	111
3.4 ÉTUDE INTRA-JOURNALIÈRE DE LA COMPLEXITÉ DES SÉRIES FINANCIÈRES : COMPARAISON DE SUPER-STATISTIQUES PAR INFÉRENCE BAYÉSIENNE	128
3.5 PREUVE D'UNE TRANSITION DE STATISTIQUES DE LA VOLATILITE STOCHASTIQUE	144
3.6 CONCLUSION	148
CONCLUSION	151
ANNEXE A	158
ANNEXE B	162
ANNEXE C	164
ANNEXE D	167
ANNEXE E	168
ANNEXE F	169
RÉFÉRENCES	174
TABLE DES FIGURES	183
LISTE DE TABLEAUX	188
TABLE DES MATIERES	190

Introduction

L'émergence de nombreuses recherches dans l'étude des systèmes complexes depuis la fin du XIX^{ème} siècle ainsi que le début du XX^{ème} siècle, notamment au travers des travaux de Jacques Hadamard (1898) [60, 65], Pierre Duhem (1906), Henri Poincaré (1908), Steve Smale (1960), Edward Norton Lorenz (1970) ou encore de Benoit Mandelbrot (1982), ont permis de mieux comprendre des phénomènes physiques à la fois présents en sciences naturelles et en sciences sociales. La notion de complexité est souvent rattachée à une volonté de définir un nouvel axe en science, permettant de révolutionner notre compréhension du comportement de certains systèmes, systèmes dont nous avons montré à ce jour notre difficulté à en prédire l'évolution ou ne serait-ce de pouvoir en avoir un certain contrôle, tel que le système cérébral humain ou encore des phénomènes sociaux ou économiques.

L'une des conséquences sine qua non de ce manque de compréhension réside dans notre incapacité à définir de manière concise ce qu'est un système complexe. A ce jour, aucun consensus n'a été établi entre scientifiques et académiciens. Ces derniers appréhendent ce problème en essayant de caractériser un système complexe au travers de ses propriétés physiques, les notions de non-linéarité, de chaos et de désordre reviennent souvent dans la littérature. Or, le risque avec cette approche, est de confondre des systèmes complexes du fait de leurs ressemblances empiriques, en termes de caractéristiques et de comportements physiques, et d'omettre le fait que probablement la ou les causes de leurs complexités respectives sont pourtant différentes. Cette tendance à assimiler ou à vouloir expliquer des phénomènes complexes par le même prisme, sans se soucier si les causes sont identiques, a fait émerger des conceptions parfois fallacieuses dans la compréhension de certain termes et théories scientifiques, comme assimiler trop souvent les systèmes complexes à des système chaotiques, ou encore assimiler l'idée de chaos à la notion d'entropie.

Ce manque de discernement, notamment dans la compréhension de systèmes sociaux économiques tels que les marchés financiers, a pu favoriser l'application de théories physiques permettant d'expliquer ou de comprendre le comportement de

systèmes complexes partageant des caractéristiques similaires mais dont les causes et fondamentaux théoriques sont bien différents. Il semble dans la littérature, notamment en sciences économiques comme dans l'étude du comportement des prix boursiers, que nous partons trop souvent du postulat que la question de la définition d'un système complexe est acquise. Par conséquent, des théories physiques telles que la théorie des systèmes dynamiques furent largement appliquées aux cours boursiers afin d'expliquer par exemple des changements de régimes de marchés, des excès de volatilité, des effets de chocs, ou encore des caractéristiques statistiques intrinsèques à savoir des distributions de loi de puissance [89].

L'un des premiers concepts qui fut largement développé dans la littérature concernant l'investigation de certain système en sciences naturelles et sociales, est le concept de marche aléatoire. C'est Louis Alphonse Bachelier qui, en 1900, formalise dans sa thèse le principe de marche aléatoire et stipule que la distribution des changements de prix boursiers suit une distribution Gaussienne [9]. Puis, quelques années plus tard, Einstein publiera la première caractérisation mathématique d'une marche aléatoire en sciences naturelles [43]. Les travaux de Bachelier ont permis l'émergence de nombreux modèles de modélisation de prix d'actifs financiers notamment le modèle Black&Scholes qui suppose que le cours des actions ou des options évolue selon un mouvement Brownien Géométrique et que les changements logarithmiques des prix sont gouvernés par une loi de distribution Gaussienne [19] à variance constante. Par conséquent, les prix des actifs financiers sont distribués selon une fonction de distribution log-normale. Il a fallu attendre le début des années 70 et notamment les travaux de Mandelbrot pour caractériser les rendements logarithmiques autrement que par le prisme d'une distribution Gaussienne, en mettant l'accent sur l'existence d'asymétries et de queue de distribution. Mandelbrot, au travers de ses travaux précédents en géométries fractales, avait réalisé que, dans l'étude des systèmes complexes, peu importe le domaine scientifique, que ce soit le domaine de l'hydrodynamique, de la mécanique des fluides, de la météorologie, de l'étude des flux de trafics routiers, de la progression de cellules cancéreuses, de la turbulence

quantique ..., il existe une caractéristique commune correspondant à l'existence d'un processus de mémoire longue, et cela sur des échelles spatio-temporelles différentes.

Pour les cours boursiers, cela se traduit entre autres par l'émergence de clusters de volatilité et d'autocorrélations dans la variation des rendements logarithmiques. Mandelbrot démontre ainsi qu'il existe un type de dépendance ou de comportement complexe qui affecte les rendements des cours boursiers [78]. Il démontre par ailleurs que cette dépendance est importante dans la variation des rendements logarithmiques absolus ou élevés au carré (i.e. la variance), en montrant la présence d'autocorrélation persistante de longue durée [39]. Les résultats de Mandelbrot suggèrent que la variance elle-même est un processus stochastique, et que c'est ce processus qui est à l'origine du comportement sous-jacent des cours boursiers. De nombreuses études corroborent les résultats de Mandelbrot, notamment des études qui démontrent l'évidence de présence de mémoire longue de la volatilité et de décroissance en loi de puissance des fonctions d'autocorrélations de la volatilité des cours boursiers [84].

Pour donner suite à ces recherches, de nombreuses études ont porté sur les propriétés statistiques sous-jacentes à la volatilité et certain formalisme ; nous pouvons les résumer en trois points :

- le degré de volatilité reflète la quantité d'information qui arrive sur les marchés
- la volatilité est l'indicateur de mesure de risque sur les marchés
- la volatilité est le paramètre indispensable à la modélisation des cours boursiers

Parmi la littérature sur l'analyse et la modélisation de la volatilité, le modèle GARCH est considéré comme le modèle de référence et fut développé dans les années 90 afin de généraliser le modèle ARCH introduit par Engle quelques années plus tôt [44]. Dans les deux modèles, chaque rendement est modélisé par la combinaison d'une variable aléatoire normalement distribuée et d'une variance locale. La variance représente la volatilité au carré ; elle varie en fonction du temps et est déterminée pour chaque instant par les variances et rendements antérieurs. En revanche, comme les modèles de variances conditionnelles ne sont pas stochastiques, Taylor décide d'ajouter

un processus aléatoire à la variance et propose un modèle de volatilité stochastique [114].

Le problème avec ce type de modèle est que la fonction d'autocorrélation de la volatilité issue de ce type de processus décroît de manière exponentielle et ne permet pas de modéliser proprement le processus de mémoire longue décrit par Mandelbrot. Afin de répondre à cette contrainte, le modèle GARCH a été étendu à un modèle FIGARCH [10]. De même, le modèle élémentaire de volatilité stochastique a été étendu à un modèle de volatilité stochastique à mémoire longue [25]. Il existe également des modèles de volatilité stochastique dont les rendements et les volatilités sont gouvernés par une équation différentielle stochastique [64].

Cependant, ces modèles de volatilité sont principalement appliqués à des séries temporelles journalières ou de fréquences relativement basses ; pourtant, l'essor des modèles quantitatifs dans l'industrie des sociétés d'investissements alternatifs nécessite l'usage de fréquence de variations de prix beaucoup plus courte, le but étant de pouvoir modéliser les cours boursiers avec plus de fiabilité [48]. La raison principale est que les données issues de séries à haute fréquence fournissent beaucoup plus d'informations sur les caractéristiques du comportement des microstructures des actifs financiers, permettant une estimation de la volatilité plus fidèle à l'observation. En effet, bien que la distribution des rendements exhibe une certaine queue de distributions à la fois sur des données hautes fréquences que journalières, les asymétries et le phénomène de concentration de volatilité sont significativement plus exacerbés sur des prix observés suivant des fréquences très petites (tic ou minute).

En outre, l'étude des données financières sur de hautes fréquences est complexe et les modèles traditionnels de modélisation de volatilité sont souvent peu adaptés puisqu'ils cherchent à maximiser la vraisemblance des données observées, mais ne permettent pas la modélisation des statistiques sous-jacentes à l'origine des comportements des actifs financiers sur ce type de fréquence.

C'est dans ce contexte qu'une nouvelle discipline a vu le jour depuis le début des années 1990, appelée Econo-physique afin de faire face aux problèmes de modélisation de la complexité associée aux marchés financiers. Le but est de sortir du prisme économétrique et d'emprunter des méthodes développées dans la recherche en physique traditionnelle sur l'étude des systèmes complexes. L'idée fondamentale n'est plus de penser en termes de modélisation de certaines caractéristiques complexes au travers de modèles économétriques et de chercher à maximiser la vraisemblance des données observées, mais, au contraire, d'essayer de trouver la distribution appropriée, autrement dit la statistique sous-jacente à l'origine du comportement de la variation logarithmique des prix des cours boursiers [108].

De nombreuses recherches font par exemple état d'analogies entre les similarités que partagent le comportement des rendements de cours boursiers et le dynamisme de certains systèmes turbulents que l'on observe en mécanique des fluides [52, 80].

Au début des années 2000, le chercheur Beck propose une analogie entre un système hors équilibre thermodynamique où la température exhibe des fluctuations [111] avec le phénomène de variance non constante des marchés financiers [18]. Il introduit alors pour la première fois un formalisme de la physique statistique appelé super-statistique qui consiste à décrire des propriétés statistiques d'un système par une superposition de statistiques. Dans le cas de la thermodynamique, on assigne une fonction de distribution à la température selon laquelle, sur de petites échelles, le système exhibe un équilibre local décrit par la statistique de Boltzmann. Sur des échelles plus grandes, le système est décrit par une superposition de la statistique de Boltzmann et de la fonction de distribution de la fluctuation de température.

L'idée fondamentale est de caractériser le système complexe de non-équilibre par une superposition de plusieurs modèles statistiques différents, indexés à différentes échelles. De nombreux systèmes complexes ont été modélisés en utilisant cette

approche de super-statistique. Y sont recensées, entre autres, l'analyse des fluctuations de vitesses de vents [73] ou encore l'analyse de la variation du niveau de la mer [95].

L'Objet de cette thèse est de déterminer et quantifier le phénomène de complexité des séries financières et d'en proposer une modélisation au moyen d'une généralisation de la statistique de Boltzmann selon des échelles de temps différentes allant de la haute à la basse fréquence.

On cherchera donc à comprendre pourquoi la variance peut être considérée comme l'information intrinsèque aux marchés, comment elle se comporte suivant les échelles spatio-temporelles considérées, et pourquoi, par son caractère prévisible, elle peut être considérée comme la conséquence sine qua none du phénomène complexe à l'origine du dynamisme des cours boursiers.

Pour répondre à cet objet, notre thèse sera structurée en trois chapitres. Dans le premier chapitre, nous nous questionnerons sur ce qu'est un système complexe ; quels en sont les caractéristiques, quels sont les pièges et les confusions à éviter. Puis, nous chercherons à en donner une définition la plus concise possible.

Dans le chapitre 2, nous ferons une revue des outils standard de mesures des phénomènes complexes. Nous introduirons entre autres les phénomènes fractales et multi-fractales et montrerons que l'approche statistique notamment au travers du prisme de la dynamique symbolique et de la théorie de l'information est une méthode fiable et rationnelle permettant de capturer au mieux la notion d'ordre et d'aléatoire. Nous montrerons également que l'un des avantages de l'approche statistique est qu'elle ne fait aucune hypothèse préalable sur la nature du comportement du système étudié. Cette dernière s'intéresse uniquement au comportement émergent des données observées.

Nous introduirons les concepts de dimension et d'entropie généralisée et montrerons que ce sont deux mesures importantes de la multi-fractalité. Puis nous proposerons une application d'analyse multi-fractale de séries financières au travers de la dynamique

symbolique et de la théorie de l'information. Nous donnerons notamment des preuves empiriques de présence de complexité et montrerons que cette dernière dépend de l'échelle spatio-temporelle d'étude.

Dans le Chapitre 3, nous introduirons au travers de la mécanique statistique de Maxwell-Boltzmann le concept de super-statistique. Nous montrerons également que les principales familles de super-statistiques sont en réalité des cas particuliers de statistiques plus générales.

Afin de considérer cette théorie comme plus appropriée pour modéliser le comportement des prix des actifs financiers sur des échelles de temps différentes, nous proposerons de modéliser ce phénomène de superposition de statistiques au travers d'expériences et de simulations numériques. En nous appuyant sur les hypothèses de la théorie des mouvements boursiers d'Osborne, nous proposons deux expériences numériques simulant la distribution des opportunités de profits d'agents financiers. Nous considérerons les rendements logarithmiques comme étant un système conservatif ou dissipatif. Ces simulations permettront de mettre en avant le concept de superposition de statistiques et de variance non constante. Pour donner suite à ces résultats, nous modifierons la théorie d'Osborne et introduirons la notion de volatilité stochastique.

Nous proposerons ensuite, par inférence Bayésienne, de modéliser, pour des échelles de temps intra-journalières différentes, le processus de la variance au travers de la statistique inverse gamma et log-normal. Nous élargirons l'étude en modélisant la distribution des rendements logarithmiques au travers du produit de la probabilité de vraisemblance des rendements avec la distribution de la variance. Nous montrerons que ces deux super-statistiques sont significatives dans la modélisation des rendements sur des échelles de temps hautes fréquences, mais perdent en significativité à mesure que l'échelle de temps s'élargit. Par conséquent, nous proposerons de prouver l'existence d'une transition de statistiques de la distribution du paramètre de variance au moyen d'une comparaison de plusieurs modèles super-statistiques sur des échelles de hautes à basses fréquences.

Nous allons maintenant revenir sur la notion de système complexe.

Chapitre 1 : Système complexe

1.1. Introduction

Les systèmes complexes ont fait l'objet depuis près de deux siècles d'études théoriques et analytiques en sciences sociales et en sciences naturelles. Si jusqu'ici aucun consensus autour d'une définition n'a pu être trouvé, de nombreux académiciens dans la littérature ont tenté d'en donner des descriptions, dont la plupart diffèrent les unes des autres.

N. Goldenfeld et L.P. Kadanoff [54] décrivent la complexité d'un système comme étant une structure subissant des variations. Cette définition permet de comprendre un premier aspect de non-stationnarité d'un système, mais reste néanmoins très abstraite.

D'autres chercheurs ont tenté de préciser la notion de complexité en distinguant selon eux une confusion faite entre :

(a) la complexité émergente d'un système dont l'évolution est très sensible aux conditions initiales ou à de faibles perturbations (système chaotiques ou système dynamique déterministe non linéaire) ayant pour conséquence l'existence d'une multiplicité de trajectoires bornées par un espace de phases dans lequel le système peut évoluer. Ces trajectoires sont régies par des équations différentielles déterministes ;

avec :

(b) la complexité émergente d'un système dont notre manque d'information à son sujet nous empêche d'en déduire un jugement ou d'en faire une description analytique.

Ce sont les chercheurs G.M Whithesides et R.F. Ismagilov [51] qui suggèrent cette contradiction existante entre une complexité émergeant de systèmes parfaitement déterministes et une autre émergeant de systèmes dont nous n'en connaissons aucune cause et qui ne semblent être gouvernés par une loi connue ou perceptible. Cette notion de mesure d'information est importante et c'est une notion fondamentale de cette thèse.

B.T. Werner [119] quant à lui associe le concept de complexité au phénomène de non-linéarité. Selon lui, la complexité d'un système est une manifestation de deux caractéristiques naturelles :

- la première émerge de trajectoires naturelles non-linéaires qui modifient ou qui se couplent avec les propriétés de l'environnement dans lequel le système évolue.
- la deuxième, qui est propre au système dissipatif, à savoir qui interagit avec son environnement comme un échange d'énergies, de matières ou encore d'informations.

Cette description met en avant non seulement la notion de non-linéarité, mais également l'aspect d'interactions entre le système et son environnement extérieur. Cette notion d'interactions revient également chez D. Rind [98] lorsque dans son article il définit le climat comme un système complexe du fait de son nombre multiple et infinitésimal d'interactions entre différents composants de l'atmosphère. En d'autres termes, ces auteurs expliquent qu'un système ne peut exhiber une certaine complexité sans cette condition d'interaction.

Richard Foote [46] part du postulat que tous les systèmes complexes partagent les mêmes caractéristiques, à savoir :

- qu'ils sont par nature complexes
- qu'ils ne sont que très rarement complètement déterministes
- qu'ils impliquent une non-linéarité ou un comportement pouvant être chaotique
- et que ces systèmes ont tendance à donner lieu à ce qu'on appelle des « comportement émergents », c'est-à-dire des comportements inattendus

Si cette caractérisation a le mérite d'être moins vague que les précédentes, cette définition reste confuse car elle fait preuve de peu de précautions dans l'usage des termes, en laissant la possibilité de confondre système non-linéaire avec imprévisibilité et chaos. Or, comme nous le montrerons dans les prochaines sections, ces termes ont un sens qui leur est propre et sont même parfois antinomiques.

Nous souhaitons mettre en évidence une définition alternative de ce qu'est un système complexe. Cette dernière fut énoncée par le physicien Nigel Goldenfield lors d'une conférence sur les systèmes complexes à l'Université de l'Illinois pendant laquelle, se basant sur ses travaux réalisés avec son confrère Leo P. Kadanoff, il lui a été demandé de s'exprimer sur la crise financière de 2008 qui a impacté le marché immobilier aux U.S.A. Il a indiqué :

« *Complexity starts when causality broken* » [54].

En d'autres termes, Nigel Goldenfield ne laisse aucune place à la possibilité de causes ou de déterminisme dans l'explication d'émergence de phénomènes complexes comme les crises financières ; cette citation va à l'encontre de la définition qu'en donne Richard Foote ou encore G.M Whithesides et R.F. Ismagilov cités ci-dessus, à savoir que les systèmes complexes peuvent être la conséquence de systèmes déterministes non-linéaires avec sensibilité aux conditions initiales et décrits par des équations différentielles.

Ceci montre à quel point ce sujet est non-consensuel. Selon Nigel Goldenfield, l'homme cherche trop souvent de manière suffisante, sous couvert d'idées trop permissives, à clamer que, derrière tout système complexe, il y a une cause ou une causalité dont notre rôle serait de la comprendre, la mesurer, l'axiomatiser et de la modéliser afin d'en contrôler le futur.

En revanche, la notion de complexité étant le fer de lance de cette thèse, nous souhaitons nous rapprocher d'une définition en énumérant les conditions nécessaires et si possible suffisantes à son émergence.

Dans ce chapitre, nous proposons, dans un premier temps, de faire une revue des caractéristiques recensées dans la littérature participant à la caractérisation des systèmes. Cela nous amènera à expliciter le lien intrinsèque qu'il existe entre système fractal et complexe ; nous expliquerons ensuite les différences entre complexité dite déterministe et dite statistique ; enfin, nous tenterons de définir ce qu'est un système complexe au travers de ses conditions nécessaires d'émergence tout en expliquant les raisons pour lesquelles ces dernières demeurent insuffisantes si elles sont prises de manière indépendante.

1.2. Propriétés et caractéristiques d'un système complexe

1.2.1. Processus non-linéaire

Il est souvent considéré comme acquis que tous les systèmes complexes partagent cette caractéristique de non-linéarité. L'aspect très intéressant de cette notion de non-linéarité réside dans l'élément de non-superposition ou de non-proportionnalité entre cause et effet. L'idée est que, considérant un système défini dans un état donné par des propriétés appelées micro-états, alors, si on modifie un micro-état parmi l'ensemble du macro-état, du fait de la divergence extrême à laquelle est soumise le système, ce changement donnera lieu à un tout autre macro-état.

C'est d'ailleurs l'une des caractéristiques les plus associées aux systèmes complexes donnant même le sentiment dans l'idée « scientifico-économique » que la non-linéarité est une condition nécessaire et indispensable à l'émergence de complexités. Pourtant, comme le décrit Robert MacKay [77], il existe des systèmes parfaitement complexes qui n'exhibent aucune dynamique non-linéaire, mais, au contraire, sont sujets à des comportements linéaires tels des modèles mathématiques modélisant des jeux d'interactions entre agents rationnels. C'est ce qui est abordé par la théorie des jeux ainsi que la théorie des systèmes en réseaux. C'est également vrai pour les dynamiques quantiques qui expliquent les mouvements et le transfert d'énergie entre systèmes. Enfin, Robert MacKay affirme également qu'il existe des systèmes linéaires complexes étudiés en sciences, notamment la théorie des matrices aléatoires

utilisée entre autres en sciences économiques et sociales dans le contexte des marchés financiers [24]. Il ajoute que les modèles stochastiques jouent un rôle prépondérant dans la description des phénomènes complexes. Pourtant, au niveau de la probabilité de distribution, les modèles stochastiques sont parfaitement linéaires, linéaires au sens que toutes combinaisons convexes (combinaisons linéaires) autorisées de distributions de probabilités, sont obtenues en faisant le produit de ces distributions. En d'autres termes, le principe de superposition est vérifié pour un système complexe si la probabilité jointe de deux variables de ce système correspond au produit de leur distribution, c'est-à-dire que ces variables sont indépendantes.

En ce sens, Robert MacKay définit la non-linéarité pour un système complexe comme étant le niveau de dépendance ou d'interdépendance entre les micro-états qui constituent ce système. Pour reprendre l'exemple de la théorie des jeux, si tous les joueurs sont indépendants, alors ce dernier est un système complexe linéaire ; dans le cas contraire où il y a dépendance entre joueurs, alors ce dernier devient non-linéaire. Dans les deux cas le système reste complexe, mais Robert MacKay souligne que la non-linéarité n'est pas une condition nécessaire à la complexité d'un système.

De la même manière, si on comprend qu'un système complexe peut exhiber des dynamiques linéaires et non-linéaires, un système dit non-linéaire n'est pas nécessairement un système complexe. En effet, si on prend l'exemple d'un pendule simple, il exhibe un comportement parfaitement non linéaire, mais il n'est en aucun cas un système complexe, tout simplement parce qu'il est très facile de prédire le mouvement de ce dernier. Pourtant le mouvement d'un pendule est caractérisé comme étant un système chaotique car sensible à des perturbations très faibles ou aux changements de ses conditions initiales.

C'est aussi une remarque que met en avant Robert MacKay dans son article. Selon lui, les systèmes complexes sont trop souvent à tort associés aux chaos, ce qui rend encore plus confuse la définition qu'il en donne. En effet, le pendule simple en est le parfait contre-exemple, à savoir qu'un système non-linéaire et chaotique n'est pas à fortiori un système complexe. S'il existe en effet des systèmes qui décrivent des complexités du fait de leur qualité chaotique, il est également tout à fait possible qu'un système

complètement chaotique décrit des trajectoires indissociables d'un comportement aléatoire. En revanche un système à double pendule est quant à lui un système à la fois non-linéaire et chaotique car sensible aux conditions initiales, mais c'est également un système complexe du fait de l'interaction et de l'interdépendance du mouvement respectif des deux pendules.

Cependant, il est important de souligner le fait que le mouvement du système double pendule est complexe, non pas du fait de la non-linéarité ou du caractère chaotique de ce dernier, mais en raison du mouvement de chaque pendule qui influe sur le mouvement de l'autre. Ces derniers créent de petites perturbations qui stimulent le caractère chaotique des pendules rendant le mouvement du système imprédictible dans sa globalité. Autrement dit, c'est l'excitation due à l'interaction du mouvement des pendules qui est à l'origine de la complexité du système. Le simple critère de dynamique non-linéaire ou de chaos n'est pas suffisant.

Ainsi, Robert MacKay postule que, dans un cadre d'étude qui se limite aux systèmes dynamiques non-linéaires, alors la définition de complexité s'adresse aux systèmes d'interactions et d'interdépendances entre agents et exclue les systèmes dynamiques de dimension basse tels que les systèmes chaotiques (c'est-à-dire avec un Lyapunov inférieur à deux [103]). De plus, puisque les comportements chaotiques sont des caractéristiques spécifiques de quelques systèmes dynamiques déterministes-et comme les systèmes dynamiques dit stochastiques peuvent aussi être des systèmes complexes et sont par définition non chaotiques, il semble que les critères de non-linéarité ainsi que de chaos ne sont des conditions ni suffisantes ni nécessaires pour définir ce qu'est un système complexe.

1.2.2. Processus auto-organisateur et auto-adaptatif

L'idée de comportement auto-organisateur est une caractéristique considérée dans la littérature comme une condition nécessaire pour expliquer l'émergence de système complexe, mais non suffisante. Ce processus d'auto-organisation ou d'autorégulation est une notion qui a émergé en théorie économique en Angleterre au

XVIIIème siècle principalement lors de la révolution industrielle. Des ingénieurs de l'époque comme James Watt (1788) cherchaient à maximiser le rendement de leur machine à vapeur en essayant d'automatiser des systèmes de régulation de pression et de transfert d'énergie. Cette idée fut ensuite théorisée par J.C. Maxwell [82]. Dans son fameux article « On Governors » publié en 1868, il développe les mathématiques pour permettre aux machines de s'adapter aux variations dues aux changements de vitesse, le but étant de garder un mouvement uniforme et de maximiser l'efficacité de production. En effet, les changements de vitesses endommageaient les machines et Maxwell a donc complexifié le système de la machine à vapeur, en rajoutant un pendule conique jouant le rôle de modérateur lorsque la vitesse de cette dernière augmente. Le mouvement du pendule interagit avec la machine, jouant ainsi le rôle de résistance permettant de ralentir cette dernière. Si le pendule est bien calibré, alors la machine doit conserver une vitesse stationnaire. Cette théorie aussi appelée théorie de contrôle est au départ une vision très mécanique dans la compréhension d'interdépendance entre agents. Dans le cas de la machine à vapeur, cela s'applique à un système non complexe, mais reste néanmoins les prémices d'une compréhension plus large d'un processus que l'on retrouve dans des comportements naturels beaucoup plus complexes notamment chez les espèces animales, végétales mais également humaines.

Les comportements auto-adaptatifs de nombreuses espèces résultent d'une interaction entre agents voisins d'un système et leurs interactions à un instant donné dépend de leurs interactions précédentes. Ces interactions sont considérées comme des échanges d'informations entre éléments (agents) d'un système plus complexe et permettent une autorégulation du comportement du système dans son ensemble. Par exemple, lorsqu'un groupe de poissons cherche à échapper à un prédateur, ils forment un amas complexe appelé « essaim » dans lequel chaque membre du groupe décide de sa position en fonction de la position et de la vitesse de ses voisins. Le but est de garder une distance constante et de conserver une forme compacte de l'ensemble du système. Si un ou plusieurs voisins viennent à changer leurs comportements (position et vitesse), chaque individu recevant cette information du fait de leurs interactions ajuste sa position.

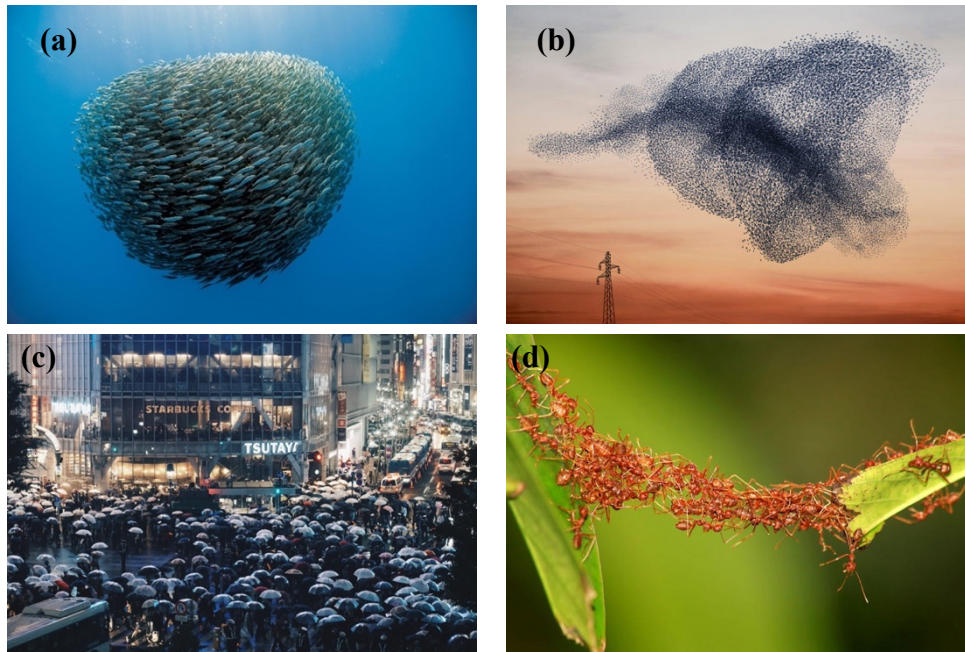


Figure 1. 1 Exemples de déplacements collectifs résultants de processus d’auto-organisation du fait d’interactions et d’interdépendance entre agents d’un système. (a) Formation d’un banc de poissons cherchant à fuir un prédateur. (b) Vol groupé d’Étourneau. (c) Croisement de piétons à Shibuya, Tokyo. (d) Construction d’un pont de fourmis servant de raccourcis aux chasseuses. (*istockphoto.com*)

Cependant, l’idée de processus auto-organisateur n’est pas suffisant pour expliquer la complexité d’un système car il nécessite un groupe d’agents important en interactions pour que ce dernier soit plausible. Si on prend l’exemple des fourmis, ces dernières sont capables de s’auto-organiser pour créer des structures complexes comme des ponts ou des habitats ordonnés, mais, prises individuellement, ces dernières sont en incapacité de reproduire de tels comportements. Les fourmis comprennent leurs tâches uniquement au travers de leurs interactions de groupe. Cet élément suppose la nécessité d’un ordre supérieur pour que le processus d’auto-organisation s’opère, peu importe que le système soit simple ou complexe.

Si on considère l’article cité ci-dessus de J.C. Maxwell, le contrôle de la vélocité de la machine à vapeur (qui est un système simple) n’est possible que par la mise en place d’un régulateur. Pour des piétons qui traversent à un croisement très fréquenté

(système complexe), leurs trajectoires s'adaptent de manière locale en fonction de leurs interactions avec leurs voisins, ce qui conditionnent et limitent leurs champs de possibilités. Avec un nombre de piétons faible, le manque d'interactions augmentera le nombre de degrés de liberté dans leurs choix de trajectoires, diminuant l'effet auto-organisateur.

L'idée centrale réside ici dans le fait que le processus d'ordre, d'auto-organisation et de contrôle est distribué et généré de manière locale pour un système complexe plutôt que de manière centrale comme c'est le cas pour la machine à vapeur.

1.2.3. *Processus ordonné et robuste*

Le processus d'auto-organisation issu des interactions d'agents n'est possible qu'au travers d'un processus supérieur qui résulte de l'émergence d'un ordre spontané. Cet ordre, selon Dimitrii Pavlov et A.O. Kasumyan [92], concernant le comportement de groupe social (dont les essaims de poissons) résulte du comportement collectif d'un système par l'agrégat d'un nombre conséquent d'interactions coordonnées entre les éléments de ce même système. Ce que l'on appelle « ordre », ce sont toutes les formes apparentes et facilement distinguables comme des symétries, des organisations précises, des structures qui se répètent, des périodicités de mouvements et de trajectoires pouvant laisser apparaître un sentiment de déterminisme. Ces structures s'engendrent par l'interactions des agents, s'auto-organisent et s'ordonnent.

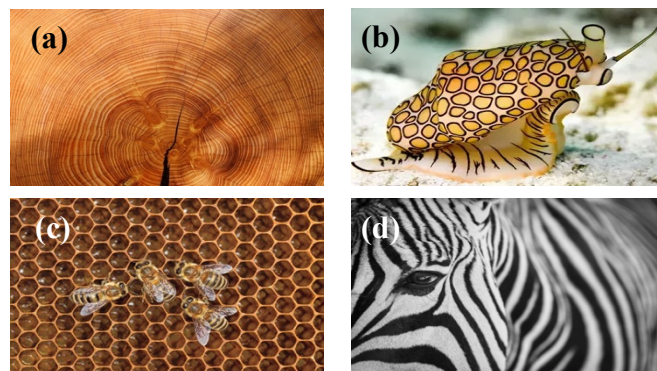


Figure 1. 2 Exemple de forme régulières symétriques et ordonnées. (1) Tronc d'arbre. (b) Escargot de mer. (c) Nid d'abeilles. (d) Zèbre. (*istockphoto.com*)

L'un des phénomènes, bien connu en biologie, relatif à l'apparition de structures organisées concerne les tâches et formes géométriques et symétriques sur le pelage de certains animaux comme les zèbres, les tigres ou encore les léopards. Ces formes ne sont en aucun cas déterminées biologiquement parlant et émergent de l'interactions entre les cellules biologiques individuelles lesquelles s'influencent les unes les autres. En émettant des substances chimiques, elles influencent l'activité chimique des autres cellules dans leur voisinage et la taille de ces formes est déterminée par la vitesse de diffusion à laquelle les cellules s'échangent des substances chimiques. Ce phénomène biologique rejoint une idée philosophique, aujourd'hui considérée comme fondamentale dans l'étude des systèmes complexes, qui est le concept d'Émergence, à savoir qu'un système pris dans son ensemble ne correspond pas à la somme de ses parties. Ce n'est pas la somme des cellules prises individuellement qui est à l'origine des formes géométriques sur le pelage des animaux, mais une propriété émergente survenue après l'interaction de ces cellules.

C'est d'ailleurs ce principe qui est à l'origine des mouvements de foule ou de panique. Les chercheurs Bayer & Rejzo en 1999, ainsi que Nilsson & Johansson [86] plus récemment en 2009, montrent, en développant des modèles de déplacements d'individus dans des espaces clos (auditoriums, théâtres...) prenant en compte l'interdépendance entre individu, que, dans des cas de panique, un individu présent dans une foule, qui verrait ses voisins s'enfuir dans une même direction, aura tendance à les imiter. L'idée de ces modèles est de montrer que lorsque l'on se trouve dans un état initial stable, l'ensemble des individus sont répartis de manière aléatoire, indépendante et uniforme dans cet espace clos. Mais, lorsque le modèle déclenche un signal d'alerte même de faible niveau (comme une alarme d'évacuation), leurs études montrent que la probabilité pour qu'un individu cesse ses activités et se dirige vers la sortie est d'autant plus importante que le nombre de personnes déjà en train d'évacuer le lieu est grand. On retrouve les mêmes caractéristiques que chez les fourmis, à savoir que, seul, un individu ne se comporterait sûrement pas de la même manière et ne reproduirait pas ce processus de mimétisme. Par ailleurs, les modèles montrent que les

endroits où le niveau de stress est le plus important sont ceux où les échanges d'informations sont maximaux avec le plus grand nombre d'individus et d'interactions.

Figure 1.3 : illustre une simulation de ce type de modèle. A l'étape 1, nous sommes en présence d'un nombre fini d'individus répartis uniformément dans un espace borné. A partir de l'étape 2 jusqu'à l'étape 6, des zones de regroupements sont visibles du fait des interactions et des effets de mimétisme. Ceci est l'origine d'émergence de formes et de structures géométriques. Nous comprenons comment il est possible pour un système complexe de se structurer ou de s'organiser autour de formes ordonnées sans en être conscient, uniquement par l'effet d'interactions et d'interdépendances avec le voisinage.

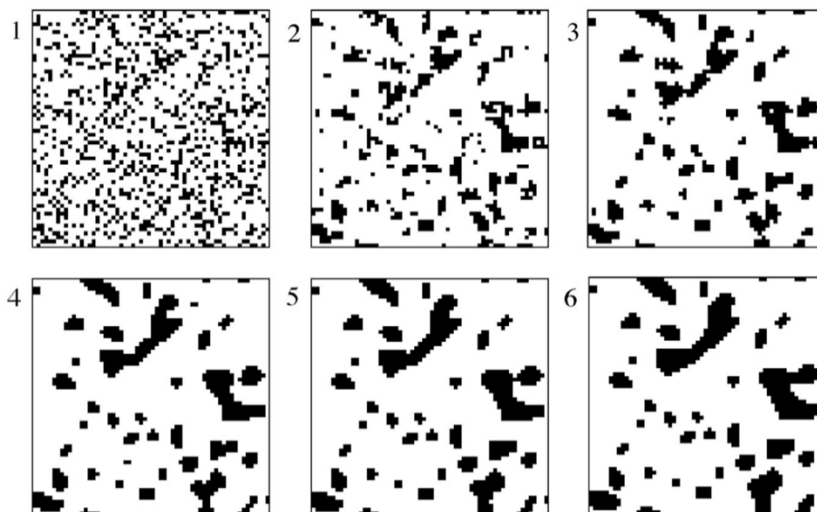


Figure 1. 3 Modèle de panique dans un espace clos et répartition des individus en zone de stress [11].

Cependant, la notion d'ordre et de formes reste très subjective et peut être uniquement basée sur des concepts purement sociétaux, ceux que l'être humain projette sur le monde physique. En revanche, le fait que les systèmes complexes fassent émerger des structures et un certain ordre spontané aussi subjectif soit-il, permet, avec un raisonnement par l'absurde, de dire qu'un système complexe ne peut pas être purement aléatoire car l'aléatoire est une condition suffisante pour l'absence d'émergence d'ordre et de structures. Et, comme indiqué ci-dessus, la condition d'interactions et

d'interdépendances entre agents est nécessaire au processus d'ordre et d'auto-organisation.

Une autre notion importante mise en avant dans la littérature concerne la notion de robustesse du système. En effet, l'ordre observé dans le travail des fourmis pour la création d'un pont, ou encore dans la manière dont les poissons s'ordonnent et se coordonnent pour former un groupe compact, est très sensible aux perturbations erratiques dues aux erreurs de mouvements ou au temps de retard dans la réception de l'information. Le système dans son ensemble reste stable peu importe les facteurs externes telles que les perturbations liées au vent ou encore aux courants d'eau.

La raison de cette stabilité est due à l'absence de contrôle centralisé du système. Les aléas se font sentir de manière locale, ne perturbant pas ainsi le système dans son intégralité et permettant, à l'échelle locale, aux agents de se ré-coordonner avec le reste du groupe [48]. En revanche, si le caractère robuste des systèmes complexes semble être nécessaire, ce dernier n'est pas suffisant car des systèmes purement aléatoires et simples peuvent être considérés comme robustes ; en effet, des perturbations n'affectent pas leur ordre puisque, par définition, ils n'en ont pas.

1.2.4. Phénomène d'Émergence et échelle spatio-temporelle

Le concept d'Émergence cité ci-dessus est une notion philosophique formalisée au XIX^{ème} siècle principalement par deux contemporains qui sont le philosophe Samuel Alexander et Conwy Lloyd Morgan dans son ouvrage « Emergent Evolution » paru en 1923. Ce dernier donne la définition de l'évolution émergente en introduisant notamment l'idée « top-down » qui est connexe au concept de hiérarchie dite de survenance. L'idée philosophique de survenance est étroitement liée au concept d'Émergence. Si on l'aborde de manière métaphysique, au sens développé par Davidson ou Bergson, la survenance est un lien de dépendance entre un contenant et ce qu'il contient, entre un état et ce qu'il survient sur celui-ci. On utilise souvent le concept de survenance pour expliquer l'interaction de l'esprit sur le corps pour justifier la dépendance ou l'impact du mental sur les états physiques. En revanche, il n'y a pas d'équivalence entre l'esprit et le corps ni de relation de causalité. Leur relation est dite

contingente dans le sens où leur interaction aurait pu être comme ne pas être ; il n'y a donc pas de cause indéterminée à leur relation ni de loi de correspondance entre les deux types d'état. Ce qui pose ainsi des limites de prédictions [85]. C'est par ce prisme de l'Émergence ou de Survenance que la science aborde la question de l'interaction des molécules et de la diffusion de substances chimiques chez les animaux zébrés comme vu ci-dessus.

Cette théorie, qui est à la fois non dualiste -car elle réfute l'idée de relation de causalité absolue entre états qui interagissent- et non réductionniste du fait de la relation de dépendance mais non d'équivalence entre états, permet d'expliquer des phénomènes évoqués ci-dessus comme l'organisation des colonies de fourmis, la formation de cristaux et, de manière plus générale, permet d'aborder l'émergence de structures organisationnelles très sophistiquées que l'on trouve dans la nature. Si le concept d'émergence permet de comprendre des phénomènes complexes, en revanche la notion d'émergence n'est pas nécessaire car des systèmes non complexes comme un gaz idéal peut exhiber des structures ordonnées.

Un point fondamental dans la notion d'Émergence ne réside pas dans cette idée de non-équivalence ou de non-causalité entre les détails d'un système (les micro-états) et le système dans son ensemble (macro-état). Il réside dans la hiérarchisation organisatrice des micro-états dans le système. La compréhension du système se fait selon différentes perspectives spatio-temporelles. L'idée est de considérer qu'il existe non seulement des dépendances entre certaines parties faisant émerger de nouvelles informations, de nouvelles structures, mais qu'il est également possible de hiérarchiser des parties, et de structurer des parties en sous-groupes, ou sous-systèmes, formant ainsi un système hiérarchisé dont certains sous-groupes sont plus importants que d'autres pour comprendre le comportement du tout. C'est ce que Herbert Simon appelle l'architecture des systèmes complexes [63]. Cette importance de présence de structures hiérarchisées dans les systèmes complexes fut approfondie par Philip Anderson [5]. Il montre qu'il est nécessaire de prendre en compte la perspective d'observations ou d'analyses du système pour comprendre l'influence et l'émergence des structures hiérarchisées et organisatrices. En effet, la compréhension d'une forêt lorsqu'on

l'observe à grande échelle sera totalement différente si on l'observe à plus petite échelle, car nous avons accès à plus de détails et plus de sous-systèmes.

Par conséquent, la compréhension d'un système dépend de l'échelle spatiale à laquelle l'observateur perçoit ce dernier. Il en est de même pour l'échelle temporelle : la description du comportement d'une forêt au Printemps sera différente de celle faite en Automne. Anderson ajoute que la structuration organisatrice et hiérarchique du système en sous-systèmes est à la fois locale et robuste (s'auto-adapte, corrige ses erreurs et est non centralisée), et que ce sont les interactions de ces sous-systèmes qui génèrent un système complexe, cela étant possible uniquement grâce à un nombre de sous-systèmes très large et qui ne cessent d'interagir.

Nous retrouvons également cette notion d'émergence en économie notamment au travers de la définition que donne Hayek¹ des systèmes économiques complexes. Selon lui, les tendances économiques comme la croissance ou l'inflation peuvent être considérées comme des phénomènes émergents parce qu'ils émanent des interactions d'agents adaptatifs et hétérogènes. Autrement dit, un phénomène économique émergent est un élément nouveau du système qui résulte des interactions d'un certain nombre d'éléments entre lesquels existent des relations simples mais qui ne peuvent être réduites aux propriétés particulières de chacun de ces éléments.

Finalement, la notion d'Émergence reste une notion philosophique difficilement abordable pour définir la complexité d'un système ; en revanche, les caractéristiques de formation de sous-systèmes hiérarchiques et organisatrices selon des échelles spatio-temporelles sont des conditions nécessaires mais non suffisantes à la caractérisation d'un système complexe. Pour affiner notre compréhension des systèmes complexes, intéressons-nous maintenant au comportement fractal de ces derniers.

¹ Hayek, F. A. (1964). The theory of complex phenomena. The critical approach to science and philosophy, 332-349. Chicago

1.3. Le comportement fractal des systèmes complexes

La géométrie fractale peut être considérée comme une extension de la géométrie euclidienne et jouer un rôle fondamental dans la description des structures hiérarchiques décrites ci-dessus.

On considère conventionnellement que, pour mesurer ou définir le degré d'une forme géométrique comme un segment, une surface ou un volume, l'utilisation d'une dimension entière comme exposant est la manière la plus simple. De ce fait, on estime qu'un segment est de dimension un, une surface de dimension deux et un volume de dimension trois. Cependant, on a conscience que toutes les formes qui nous entourent ne présentent pas ce type de régularités, et qu'il est très facile d'apprécier des formes beaucoup plus complexes comme la courbe de Peano ou encore le tapis de Sierpinski (figure 1.4). De même, la côte océanique de la France n'est pas une ligne parfaite, ni une surface, c'est une forme plus complexe dont la dimension est quelque part entre 1 et 2 [106]. La géométrie Euclidienne ou géométrie non-fractale (un point unique, un segment, un cercle, un cube...) pourrait être considérée comme des objets fractals avec le niveau de complexité le plus faible possible (une dimension fractale en entière).

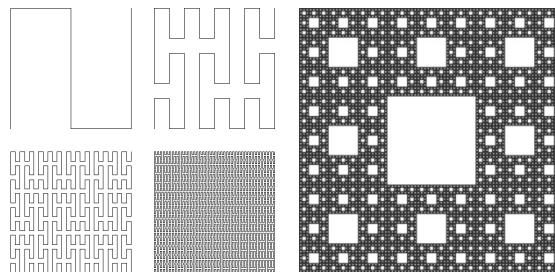


Figure 1. 4

(a) Courbe de Peano.

(https://en.wikipedia.org/wiki/Peano_curve)

(b) Tapis de Sierpinski.

(https://en.wikipedia.org/wiki/Sierpi%C5%84ski_carpet)

Le terme originel « fractal » vient du latin *fractus*, signifie « irrégularité » ; il fut utilisé pour décrire des systèmes spatio-temporels (des objets dans l'espace ou des variations dans le temps) qui partagent des similarités ou des autosimilarités qui ne peuvent pas être décrites et analysées via une unique échelle de mesures. Les fractales

sont donc des formes géométriques irrégulières en espace et en temps, caractérisées par des motifs qui se répètent sur des échelles différentes (échelles spatiale et temporelle). Cette géométrie fractale implique qu'une portion d'un objet fractal ou d'une série possède les mêmes caractéristiques statistiques de l'ensemble de l'objet. La géométrie fractale est aujourd'hui considérée comme une science à part entière, capable d'expliquer certains phénomènes issus de systèmes complexes, à savoir la capacité de ces systèmes à s'auto-organiser suivant des formes et motifs imbriqués sur des échelles d'espaces ou de temps différents.

Nous pouvons mesurer le degré de fractalité d'un système au moyen de ce que nous avons évoqué plus haut, à savoir la dimension fractale D , qui permet de mesurer le degré d'ajout de détails de structures lorsque l'échelle d'observation augmente. Il existe plusieurs types de Dimensions fractales qui dépendent principalement de la nature du système étudié ; ces dernières seront explicitées dans le chapitre 2. Mais, de manière générale, la dimension fractale est une mesure de rugosité d'un système ; plus sa valeur est grande, plus le système peut être considéré comme rugueux. Par exemple, le tapis de Sierpinski (Figure 4) se caractérise par une dimension fractale de 1.58, la dimension du cerveau humain est comprise entre 2.73 – 2.79 et celle des membranes alvéolaires d'un être humain équivaut à 2.7 [69]. C'est en ce sens que W. Klonowski définit la dimension fractale comme mesure de complexité [69], et nous reviendrons sur cette notion dans le chapitre 2.

Tous les objets fractals connus et étudiés au travers de systèmes et modèles physiques ont deux types de propriétés qui permettent de définir deux classes de fractals :

- Les fractals dits mono-fractals, qui sont statistiquement auto-similaires, se caractérisent par une unique dimension. On dit que ces fractales sont self-similaires, invariant aux changements d'échelles, et conservent la même dimension, peu importe l'échelle d'observation. La similarité ne vient pas de l'exactitude des formes et des structures sur des échelles différentes, mais du fait que ces dernières obéissent aux mêmes statistiques. Cependant, la notion d'invariance ou de symétrie ne peut être une notion générique pour la simple

raison que tous systèmes réels et observables interagissent avec leur environnement et leur voisinage et sont donc soumis à des perturbateurs qui peuvent détruire l'aspect symétrique et invariant de ces derniers.

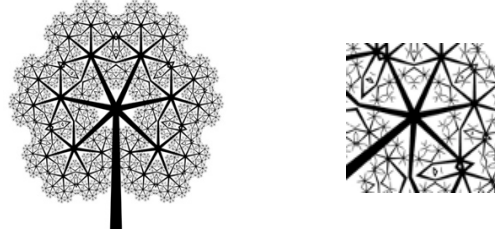


Figure 1. 5 Self-similaire fractales. Similaire apparence sur des échelles plus petites. La structure de droite est similaire à l'ensemble de la structure représentée à gauche. (<https://www.rgbstock.com/photo/omq8JIC/Fractal+Tree+8>)

- Les fractales dites « multi-fractales », possédant des dimensions différentes selon les échelles d'observation. On dit que ces multi-fractales sont self-affines ; elles se constituent d'un ensemble de fractales qui se superposent sur des échelles différentes. Elles se caractérisent par un spectre infini de dimension généralisé, aussi appelé le spectre de dimension de Rényi [66]. Nous reviendrons sur cette notion de spectre de dimension dans le Chapitre 2.

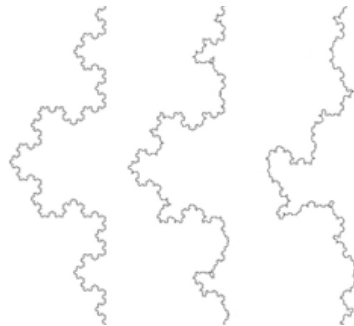


Figure 1. 6 Construction itérative de courbes fractales dites self-affine dont leur dimension respective dépend du facteur d'échelle d'observations. (<https://www.rgbstock.com/photo/omq8JIC/Fractal+Tree+8>)

Ainsi, la différence fondamentale entre la famille de fractales self-similaires et self-affines réside dans le fait que les changements d'échelles produisent ou non des équivalences statistiques. Pour les fractales self-similaires, il suffit de changer d'échelle avec le même facteur dans toutes les directions pour pouvoir conserver des structures avec des statistiques identiques. Alors que, pour les fractales self-affines, garder le même ratio d'échelles ne suffit pas. Il faut adapter et modifier le ratio de changement d'échelles en fonction des différentes directions possibles pour pouvoir produire des structures avec des statistiques équivalentes. En d'autres termes, une structure self-affine est statistiquement équivalente selon l'échelle d'observation sous condition de subir une transformation affine [41] .

Le concept de géométrie fractale est né dans le domaine de l'économie lorsque B. Mandelbrot ² a étudié la fluctuation des prix de marchés de produits dérivés. Il a trouvé de manière empirique que la variation des changements de prix du marché du coton selon des échelles de temps différentes était similaire. A la suite de ce constat, Mandelbrot a compris qu'une telle invariance d'échelle pourrait être la clé pour caractériser de nombreux phénomènes complexes, notamment les systèmes économiques. Il est important de souligner que les propriétés fractales des prix des marchés financiers sont de plus en plus exacerbées depuis le début des années 1990 ; cela s'explique par l'informatisation des transactions financières et la démocratisation du « trading » aux particuliers, stimulant l'aspect hétérogène des agents financiers. Cette découverte a permis l'essor de nouveaux domaines scientifiques comme « l'éconophysique ³ », une étude des phénomènes économiques à partir des méthodes et approches physiques, en particulier les modèles physiques des distributions de loi de puissance. Parmi les nombreux sujets en « éconophysique », il y a trois sujets qui sont

² B.B. Mandelbrot, *Int. Eco. Rev.* 1 (1960) 79, B.B. Mandelbrot, 'The Fractal Geometry of Nature' (Freeman, San Francisco, 1983).

³ J. Kertesz and I. Kondor (Eds.), *Econophysics : an emerging science* (Kluwer Academic Publisher, Dordrecht, to appear).
R. Mantegna and E. H. Stanley, *Introduction to Econophysics* (Cambridge Univ. Press, London, to appear).

étroitement liés à l'étude des fractales. Ce sont les variations de prix sur les marchés financiers, la répartition des revenus dans l'économie et la relation d'échelles des fluctuations de la taille de l'entreprise.

1.4. Différences entre complexité statistique et complexité déterministe

L'une des conceptions largement rependues sur les systèmes complexes est de considérer que la complexité se situe entre deux extrêmes, l'ordre parfait et le hasard certain.

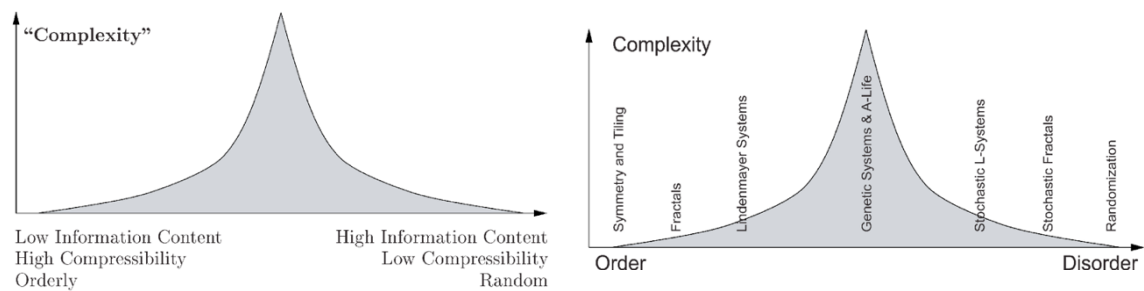


Figure 1. 7 Figure provenant de “The Computational Beauty of Nature: Computer Exploration of Fractals, Chaos, Complex Systems and Adaptation”. Copyright 1998-2000 by Gary William Flake.

Les systèmes complexes ne sont jamais parfaitement ordonnés mais ne sont également jamais totalement aléatoires dans le sens de structures ou d'ordres apparents. C'est d'ailleurs cette dernière qui a entre autres participé aux nombreuses recherches liées aux systèmes chaotiques dans l'étude des systèmes complexes du fait de leurs caractéristiques imprévisibles proches des comportements aléatoires, tout en ayant pourtant un ordre caché du fait de leur caractère déterministe. Ceci est à l'origine de confusions dans les termes et à la relation, certes réelle entre chaos et complexité, mais qui n'est ni nécessaire ni suffisante.

Les systèmes chaotiques peuvent être plus ou moins définis de manière stricte ; il existe également différentes approches. Nous retiendrons celle de Paul Davidovits [38] donné dans le recueil « Physics in Biology and Medicine (Fifth Edition) », laquelle, dans le contexte de la science, définit un système chaotique par son caractère déterministe et de sensibilité extrême à ses conditions initiales. Un système

déterministe est chaotique si ses trajectoires divergent exponentiellement au cours du temps du fait de variations de ses conditions initiales ou du fait de perturbations.

Si un système complexe peut être de cet ordre, comme le double pendule par exemple, un système complexe à proprement dit ne partage pas forcément ce type de propriétés ni ce type de comportement. Un système complexe est nécessairement défini par un nombre de degrés de liberté très important, constitués de sous-ensembles qui interagissent et qui sont partiellement dépendants. Et, selon le chercheur Yanner Bar-Yam [11], dans son ouvrage « Dynamics of Complex Systems », il n'est pas évident que cette interaction soit toujours modélisable par les systèmes chaotiques du simple fait qu'ils soient conditionnés par des équations déterministes qui décrivent principalement des systèmes fermés et non dissipatifs.

Par conséquent, le chaos concerne le dynamisme issu de la variation de quelques paramètres, alors que les systèmes complexes -pour atteindre un certain niveau de dynamisme- doivent interagir avec un grand nombre de variables présentes dans leur voisinage. De plus, Yann Bar-Yam ajoute que la notion de chaos est souvent utilisée pour faire référence au caractère désordonné ou imprévisible des systèmes complexes ; il ajoute que les notions de désordre ou d'aléatoire sont très distinctes des concepts de chaos déterministes. Ainsi, il préconise que leur assimilation doit être exclue pour éviter toutes confusions. C'est pourquoi il préfère remplacer la locution « environnement chaotique » par « environnement complexe ».

Ainsi, contrairement au chaos, les systèmes complexes traitent à la fois de la structure de la dynamique des systèmes et de leurs interactions avec leur environnement. De plus, si les systèmes chaotiques font émerger des comportements situés entre l'ordre et le désordre, des systèmes stochastiques peuvent parfaitement exhiber ce type de caractéristiques bien qu'ils ne soient certainement pas régis par des équations déterministes. Si on prend comme exemple la méthodologie utilisée par Shannon en 1943 [101] pour analyser la structure d'un message, on comprend que des langues comme l'Anglais ou le Chinois sont des systèmes complexes dans la mesure où ils aboutissent à une série de symboles (alphabets que Shannon appelle événements) qui

se caractérisent par un système stochastique ni déterministe (le symbole suivant ne peut être calculé ou déterminé avec certitude) ni aléatoire (le symbole suivant n'est pas totalement libre). Les langues sont régies par un ensemble de probabilités ; chaque symbole ayant une probabilité qui dépend de l'état du système actuel et de ses états antérieurs.

Par conséquent, il est essentiel d'éviter de confondre le chaos avec des systèmes complexes, car ce sont des notions certes liées mais distinctes par nature. Si le Chaos implique l'émergence de complexités, les systèmes complexes ne sont pas nécessairement régis par des systèmes chaotiques.

L'approche statistique quant à elle considère qu'un système complexe se définit par sa capacité organisationnelle qui persiste dans le temps et ne fait aucun postulat sur l'origine du comportement complexe. Il y a deux approches concernant l'alternative statistique.

La première concerne l'approche de Bennett [18] qui s'inspire grandement des travaux de Shannon sur la Théorie de l'information. Il formalise l'idée que les systèmes complexes sont ceux dont les explications les plus plausibles décrivent de longs processus causaux. Bennett cherche à définir un système complexe au travers d'un ensemble de données historiques causales auxquelles sont attribuées des probabilités de plausibilité. C'est ce que Bennet appelle « La profondeur Logique ». Ce concept considère que le comportement d'un système complexe se comporte comme un algorithme prédéterminé avec une grande profondeur logique, trop grande pour pouvoir vraiment en comprendre les structures émergentes. En revanche, Crutchfield et Shalizi [33], souligne que l'approche de Bennet ne donne aucune information sur la longueur des trajectoires qui doivent être considérées. Si la complexité est une propriété d'un processus, comment identifier l'origine du processus ? C'est pourquoi Crutchfield et Shalizi proposent d'examiner la vitesse à laquelle la profondeur augmente, plutôt que sa valeur absolue. Or, le taux de croissance de la profondeur s'avère être le taux avec lequel l'entropie du système augmente, ce qui revient à la mesure d'Information de Shannon.

L'approche de Shannon est la deuxième alternative statistique. L'idée est de considérer un ensemble de données se caractérisant par un certain nombre de symétries, tout en conservant son processus aléatoire. Plus tard, Grassberger [57] propose de définir la complexité d'une série de données convertie en série de symboles (un alphabet ou des bits) comme étant la quantité moyenne par laquelle l'incertitude d'un symbole dans un sous-ensemble de données diminue en raison de la connaissance des symboles précédents. Pour une série de nombres complètement aléatoires, cette quantité est nulle car aucune information n'est obtenue par la connaissance d'autres symboles aléatoires (du fait de leur non-dépendance). Plus il y a de dépendances statistiques, c'est-à-dire d'ordre, plus la complexité augmente et la mesure de cette quantité résulte de l'entropie de Shannon.

Ainsi l'approche statistique s'intéresse à l'émergence d'ordres issus d'un processus dont on ne connaît pas la qualité déterministe ou aléatoire. Alors que l'approche déterministe s'intéresse à l'émergence du désordre issu d'un processus considéré comme étant parfaitement déterminé.

Cette idée d'incertitude est également une composante importante lorsque l'on souhaite caractériser l'économie de systèmes complexes. L'économiste Arthur ⁴postule que l'état de non-équilibre est l'état naturel de l'économie. L'économie est toujours dans un état de flux, un état dissipatif, en constante évolution et changement. Selon Arthur, il y a deux raisons principales à cela : l'une est l'incertitude fondamentale, l'autre l'innovation technologique. Le concept d'incertitude fondamentale en économie fut introduit par Knight ⁵et Keynes ⁶en estimant qu'une distinction devait être faite entre risque et incertitude. En cas de risque, tous les événements futurs possibles sont connus et nous pouvons en calculer la probabilité d'occurrence. Cependant, il existe de nombreuses situations dans lesquelles nous ne connaissons pas toutes les issues

⁴ Arthur, W. B. (2013) Complexity Economics: A Different Framework for Economic Thought.

⁵ Knight, F. H. (1921): Risk, Uncertainty and Profit. Boston: Houghton Mifflin

⁶ Keynes, J.M. (1921), A Treatise on Probability, The Collected Writings of John Maynard Keynes, Vol. VIII, London

possibles. Dans ces situations d'incertitude, le calcul de probabilités dit « Frequentist » devient moins adapté et l'inférence bayésienne prend plus de sens. Ainsi, pour faire face à cette incertitude, les agents économiques essaient de donner un sens aux problèmes en faisant des suppositions, en utilisant les connaissances et expériences passées et en utilisant des modèles heuristiques de décisions simples comme des règles empiriques⁷. En conséquence, les agents mettent continuellement à jour leur modèle de prise de décision interne, ce qui signifie qu'ils adaptent, rejettent et remplacent constamment les actions ou les stratégies en fonction de leurs expériences. De telles dynamiques sont connues sous le nom de dynamiques évolutives, une branche de la théorie des jeux, développée par Lindgren ⁸ permet de démontrer qu'en présence de bruits, les modèles évolutionnaires ne se retrouvent généralement pas dans un équilibre de Nash stable. Si ces résultats peuvent être extrapolés à l'ensemble de l'économie, ils doivent être en permanence en mouvement perturbateur au fur et à mesure que les agents explorent, apprennent et s'adaptent à leur environnement.

⁷ Gigerenzer & Gaissmaier 2011, Heuristic Decision Making, Annual Review of Psychology, Vol. 52:451-482, 2011

⁸ Lindgren, K. (1997). Evolutionary dynamics in game-theoretic models. In *The economy as an evolving complex system II* (pp. 337-368).

1.5.Conclusion

Ce chapitre propose une définition de ce qu'est un système complexe au travers d'une revue des caractéristiques physiques. Nous essayons également d'éliminer certaines confusions, notamment en expliquant les distinctions entre complexités statistique et déterministe. On en a déduit une série de conditions nécessaires à la caractérisation d'un système complexe et proposé une définition.

Nous pouvons définir la notion de complexité par le biais de quatre notions qui sont :

- **Un ensemble constitué de nombreux sous-ensembles** : pour que les interactions se produisent, les éléments de cet ensemble doivent être non seulement nombreux, mais également de nature similaire dans leur objectif final. Dans l'exemple de l'essaim de poissons, lorsque des individus sont menacés par un prédateur, leur objectif final commun est la survie. Pour que les éléments du système puissent interagir ou communiquer, ils doivent pouvoir échanger de l'énergie, de la matière ou de l'information. Les éléments doivent être comparables et soumis aux mêmes lois physiques ou sociales. En biologie, les cellules, avant de former des organismes multicellulaires, sont indiscernables voire identiques, de sorte qu'elles peuvent communiquer au maximum (échanger de l'information génétique) les unes avec les autres. En revanche, s'il est nécessaire pour un système complexe d'être constitué de nombreuses parties similaires d'un certain type, il convient de noter que toutes les parties ne doivent pas nécessairement être similaires et du même type. Les systèmes météorologiques sont composés de nombreuses parties similaires puisque l'oxygène, l'azote et l'eau sont tous des gaz composés de molécules, mais ces gaz ont très différents dans leurs propriétés et ces différences donnent lieu à des caractéristiques importantes du climat. C'est aussi vrai pour des systèmes non physiques comme les structures sociales du type marchés financiers qui sont constitués d'agents avec des caractéristiques d'investissements différentes.

En revanche, ces agents financiers sont soumis à des règles similaires et partagent une volonté commune qui est de maximiser leur opportunité de profit sur le long

terme. Et l'interaction entre ces agents correspond à l'information émise par le marché. Nous verrons plus en détail cette notion dans le chapitre 3.

- **Présence d'interactions** : La seconde condition est que les éléments du système aient les moyens d'interagir. L'interaction peut être un échange d'énergie, de matière ou d'informations. Le mécanisme de médiation peut être des forces, une collision ou une communication. Sans interaction, un système ne forme qu'un amas d'éléments qui sont nécessairement indépendants et n'ont aucun moyen de former des structures, des motifs ni d'établir un certain ordre. Ici, nous devons souligner que l'interaction doit être directe, et non via un tiers ou une cause commune.

Nous exigeons non seulement une dépendance probabiliste, mais encore une dépendance directe entre éléments. Par ailleurs, la localité d'interaction n'est pas une condition nécessaire ; elle n'est pas non plus suffisante. Les interactions peuvent être canalisées par des systèmes de communication et de transport spécialisés.

Par exemple, les lignes téléphoniques ou les wifi permettent des interactions non locales entre les agents d'un marché financier même sur des distances très grandes.

De ce fait, il est important que l'idée d'interaction ne soit pas ici nécessairement une question de dynamique. Ce qui est important, c'est l'idée de la dépendance des états des éléments les uns par rapport aux autres.

- **Désordre** : les systèmes complexes sont précisément ceux dont l'ordre émerge du désordre. L'ordre ne se construit pas à partir du désordre, de sorte que le désordre devienne plus ordonné, mais émerge de celui-ci et s'ajoute au désordre présent. L'interaction est à la base de toute corrélation qui se construit et donc à l'origine de l'ordre qui émerge du désordre. D'où l'intérêt que l'on porte dans cette thèse à l'approche statistique dans l'étude des systèmes complexes.

Un exemple d'ordre émergent du désordre peut être donné avec les fluctuations thermiques qui sont nécessaires pour que la plupart des processus biologiques aient lieu, comme la liaison des protéines à l'ADN qui est entraînée par l'énergie fournie par les

fluctuations thermiques. Le résultat est la transcription de l'ADN pour fabriquer de nouvelles protéines.

- **Un ordre robuste** : Les trois conditions ci-dessus sont toutes nécessaires pour qu'un système complexe émerge, mais elles ne sont pas suffisantes. En effet, comme évoqué ci-dessus, de nombreux éléments similaires interagissent de manière désordonnée dans un gaz ; pour autant, ce ne sont pas des systèmes complexes. Lorsque les éléments continuent d'interagir de manière désordonnée, il faut également que les structures globales soient préservées. Le niveau macroscopique de l'ensemble résultant d'une interaction microscopique doit être stable sur une échelle spatio-temporelle considérée. Ce type d'ordre robuste est une autre condition nécessaire pour qu'un système soit complexe. Par ailleurs, un corollaire direct de l'ordre robuste est la notion de mémoire, de sorte que le système garde une trace de passé grâce à la persistance de sa structure interne.

On propose ainsi une définition plus générale par l'énoncé suivant :

Un système complexe est un ensemble dynamique constitué de sous-ensembles d'éléments qui interagissent d'une manière aléatoire, désordonné et imprévisible, faisant émerger au niveau de l'ensemble des structures ordonnées et hiérarchisées. Grâce à un système de contrôle et d'apprentissage localisé, ce dernier est capable de maintenir ses structures tout en interagissant avec un environnement non stationnaire.

Cette définition constitue l'hypothèse à l'origine de notre choix d'étudier et d'analyser la complexité des séries financières économiques au travers d'une approche statistique et selon des échelles spatio-temporelles différentes.

Après avoir défini la notion de complexité, nous allons nous intéresser dans un deuxième temps à la mesure de cette dernière au travers d'une approche statistique.

Chapitre 2 : Mesure de la complexité : une approche statistique

2.1 Introduction

Dans ce Chapitre, nous nous intéressons à la notion de mesure des systèmes complexes. Il y a deux approches de mesures communément utilisées. La première s'applique aux données qui sont obtenues d'un système quelconque et la deuxième s'applique directement au système lui-même.

Les mesures de complexités qui s'appliquent au système font références à la théorie de Profondeur de Bennett, la théorie de Profondeur thermodynamique ou encore à la mesure de complexité effective définie en 1996 par Murray Gell-Mann et Seth Lloyd [74, 75, 49, 50].

En revanche, celles qui s'appliquent directement aux données font référence aux approches statistiques, mesures introduites par James Crutchfield et Karl Young [31, 32, 34] qui traitent une manière effective de mesurer le traitement de l'information des systèmes dynamiques non linéaires.

Cette approche statistique est une technique de reconstruction de l'état causal, c'est-à-dire qu'à partir de données empiriques ou d'une description probabiliste du comportement du système, on déduit un modèle du processus caché qui génère le comportement observé. L'idée de l'inférence statistique est que sont recueillies des probabilités pour des observations successives conditionnées à des observations antérieures-, à partir des données empiriques. De tels chemins possibles (séquences de types d'observation) sont placés dans la même classe d'équivalence lorsqu'ils ont la même probabilité conditionnelle pour des observations futures. Les ensembles

résultants définissent chacun des prédictions pour les observations futures et, par conséquent, chacun constitue un certain niveau de régularité du système.

Ainsi, une mesure statistique de la complexité cherche à quantifier le niveau de causalité entre états d'un système et, connaissant ses états précédents, représente notre capacité à prédire le comportement futur du système. La complexité statistique contient toutes les informations disponibles sur l'organisation du processus d'un système. Un système est donc statistiquement complexe s'il présente un grand nombre d'états causaux et donc un grand nombre de régularités.

Nous nous intéresserons dans ce chapitre à une mesure de la complexité au moyen d'une approche statistique faisant état des données recueillies du système étudié. La raison principale qui nous pousse à préférer l'approche statistique est liée à la définition des système complexes donnée dans le chapitre 1, dans laquelle nous mettons l'accent sur la notion d'émergence d'ordre d'un processus initialement non ordonné. Nous soutenons l'hypothèse qu'une bonne mesure de la complexité est une mesure qui n'est pas maximale lorsque le système est parfaitement aléatoire (ce qui serait le cas pour une mesure dite déterministe), mais au contraire une mesure qui permet de quantifier le niveau de création d'ordre, le degré d'organisation et de régularité d'un système avec une exigence : que cette dernière soit calculable.

Ainsi, l'approche statistique est la méthode qui saisit le mieux la notion qualitative de l'ordre produit par des système complexes et elle est parfaitement calculable car elle correspond à la mesure du taux de changement d'entropie de Shannon du système étudié.

Ce chapitre s'articule autour de quatre sections. La première met l'accent sur une mesure géométrique de la complexité d'un système au travers du calcul d'une dimension fractale. La deuxième section introduit une mesure statistique fondée sur la théorie de l'information de Shannon et de la dynamique symbolique. Nous démontrons que mesurer le niveau d'organisation revient, selon Shannon, à quantifier l'information qui émerge d'un système. La troisième section généralise l'entropie et la dimension de

Shannon, notamment via le spectre de singularité et l'entropie de Rényi. La quatrième section applique ces méthodes d'analyse et de mesure de complexité à une série temporelle de rendements logarithmiques de l'indice boursier Américain : le Standard&Poor500. On montre la dépendance du caractère complexe et prédictible de ce dernier aux échelles spatiotemporelles d'études.

2.2 Les Dimensions Fractales

Comprendre le dynamisme des systèmes complexes, leur nature et pouvoir les quantifier au travers de mesures fiables et descriptives, est devenu essentiel pour de nombreuses disciplines scientifiques telles que l'étude de l'activité solaire [70], le diagnostic de maladies de type pathologies cardiaques [41, 7, 71, 90, 94, 115], la détection de tumeurs [110], l'étude de la variabilité de la température atmosphérique [55], l'étude de phénomènes de sociétés tels que le dynamisme des décisions présent dans des conseils d'administration [12], et bien d'autres.

C'est dans cet objectif que l'analyse et la quantification des fractales sont devenues depuis le demi-siècle dernier un sujet d'étude très important, motivé par le fait que tous les systèmes complexes -peu importe l'origine de leur complexité- se caractérisent par l'émergence de formes géométriques hiérarchisées et auto-organisatrices que l'on appelle des structures fractales.

L'un des outils le plus important dans l'étude des structures fractales correspond à la dimension fractale. Cette dernière permet de quantifier le degré d'information qu'offre un système sur son niveau de complexité. Autrement dit, la dimension fractale d'un système permet de déterminer le degré de liberté auquel est soumis le système étudié. Il existe plusieurs définitions et méthodes de calculs concernant la dimension fractale dont deux sont plus communément utilisées à savoir la dimension de Hausdorff définie en 1919 [123] et la dimension de capacité aussi appelée « box-counting » [107]. Plus récemment des méthodes statistiques alternatives, telles que la dynamique symbolique, furent introduites, notamment grâce aux travaux de Shannon sur la théorie de l'information.

Dans la suite de cette section, nous donnons les définitions de dimensions fractales les plus fréquemment utilisées et nous montrons que ces dimensions sont toutes des cas particuliers d'une définition de la dimension plus générale que l'on appelle dimension généralisée ou dimension de Rényi. Nous verrons d'ailleurs, dans le cas de systèmes complexes avec facteur d'échelle non-invariant tel que les multi-fractales, que l'approche généralisée doit être privilégiée.

2.2.1 Dimension et mesure de Hausdorff

Felix Hausdorff, mathématicien Allemand considéré comme l'un des pionniers et fondateurs de la topologie moderne est également à l'origine du terme « Dimension fractale ». Ce terme vient de ses travaux notamment sur la mesure d'Hausdorff [123] correspondant à une généralisation des notions de mesures traditionnelles, telles que l'aire d'un carré ou le volume d'un cube qui sont des dimensions entières, à des mesures plus complexes s'adressant à des formes géométriques complexes (comme les fractales). Ces formes géométriques ne sont ni complètement des carrés, ni complètement des cubes, mais des formes intermédiaires qui nécessitent des mesures métriques non-entières.

Bien que la définition rigoureuse de la dimension fractale d'Hausdorff ne permette pas de calculer immédiatement la valeur de cette dernière, elle permet néanmoins de fournir le cadre théorique sur lequel des algorithmes d'estimations de dimension peuvent s'appuyer. La dimension d'Hausdorff est uniquement une description de la géométrie fractale d'un objet ou d'un système physique ; mais elle ne permet pas de décrire explicitement une moyenne statistique μ de la fractale pour la raison évidente que cette mesure statistique dépend de l'échelle ε d'observation selon une loi de puissance telle que :

$$\mu_2/\mu_1 = (\varepsilon_2/\varepsilon_1)^h \quad (2.1),$$

avec h l'exposant d'échelle.

De ce fait, Hausdorff commence par définir sa propre mesure Γ qui correspond à une distribution de densité uniforme sur l'ensemble fractal. Il appelle \mathcal{A} l'ensemble fractal dont nous souhaitons calculer la dimension. Il appelle $\mathcal{C}(r, \mathcal{A}) = \{\mathfrak{B}_1, \mathfrak{B}_2, \dots, \mathfrak{B}_k\}$ un ensemble fini qui recouvre l'ensemble \mathcal{A} au moyen d'une réunion dénombrable de parties \mathfrak{B}_k , chacune étant de diamètre inférieur à r . Par conséquent, on note l'ensemble des parties $\mathcal{A} \subset \bigcup_{i=1}^k \mathfrak{B}_i$ dont le diamètre de chacune d'entre elles satisfait $\delta_i = \delta(\mathfrak{B}_i) < r$. L'idée de contraindre le diamètre de chaque partie au travers d'une valeur de majoration r est de volontairement minimiser le rôle des parties \mathfrak{B}_i lors du recouvrement de la fractale dans le calcul de la dimension de \mathcal{A} .

Ainsi Hausdorff considère la fonction suivante :

$$\Gamma(\mathcal{A}, D, r) = \inf_{\mathcal{C}(r, \mathcal{A})} \sum_i \delta_i^D \quad (2.2),$$

où *inf* est le infimum (minimum) de l'ensemble des parties de recouvrement de $\mathcal{C}(r, \mathcal{A})$ sur \mathcal{A} qui respectent la contrainte de majoration $\delta_i < r$.

Cet infimum représente une mesure de granularité (« coarse-grained ») de l'ensemble \mathcal{A} . Par exemple, si $D = 1$, alors $\Gamma(\mathcal{A}, D, r)$ donne la longueur de l'ensemble \mathcal{A} avec une règle de mesure de longueur r , et plus r tend vers 0, plus Γ se rapproche de la longueur réelle de \mathcal{A} . On comprend aisément que, pour une valeur de r très petite de telle sorte que $r \rightarrow 0$ et pour la plupart des valeurs choisies de D , la valeur de Γ divergera vers l'infini (que $\Gamma \rightarrow \infty$) ou convergera vers 0 ($\Gamma \rightarrow 0$). Si on imagine une figure qui se caractérise par un segment de longueur finie, alors dans ce cas elle aura une aire égale à zéro et, dans le cas où on imagine une figure qui se caractérise par une aire finie, alors elle sera recouverte par une infinité de petites longueurs.

Enfin, puisque $\Gamma(\mathcal{A}, D, r)$ est une fonction qui décroît de manière monotone en fonction de D , cela implique l'existence d'une limite (éventuellement infinie) quand on fait tendre r vers 0. D'où, l'existence d'une transition unique vers la définition de la dimension de Hausdorff notée D_H telle que :

$$\Gamma(\mathcal{A}, D) = \limsup_{r \rightarrow 0} \Gamma(\mathcal{A}, D, r) = \begin{cases} \infty & \text{for } D < D_H, \\ 0 & \text{for } D > D_H \end{cases} \quad (2.3),$$

avec $D_H = \inf\{D: \Gamma(\mathcal{A}, D) = 0\}$ qui définit la dimension de Hausdorff.

Par exemple, la fractale d'un littoral pourrait être recouverte par une infinité de parties de dimension 1 (un segment) et une infinité de volumes de dimension 2 (une aire). Ainsi, ce littoral sera défini par une dimension D comprise entre 1 et 2. Par ailleurs, la mesure de granularité donnée par l'équation (2.2) exhibe le facteur d'échelle (le changement d'échelle de la fractale) tel que $\Gamma(\mathcal{A}, D, r) \sim r^{D - D_H}$, ce qui donne une autre manière de définir la dimension de Hausdorff sans que la valeur de r soit nécessairement très petite.



Figure 2. 1 Exemple d'une mise à l'échelle d'un système par recouvrement de ce dernier avec des formes géométriques de faible diamètre ou de faible longueur (cercle, carré) pour déterminer sa dimension avec la méthode de Hausdorff. (google.com/images.com)

La difficulté majeure, avec l'utilisation de la dimension de Hausdorff comme mesure de la complexité d'un système, résulte de l'implémentation numérique du fait de la valeur infimum qui nécessite un nombre de calculs important, notamment pour des systèmes très rugueux. Pour contourner ce problème, on utilise une autre définition de la dimension qui résulte de la dimension de Hausdorff appelée « dimension de capacité ».

2.2.2 Dimension de Capacité (Box Counting)

Comme expliqué précédemment, plus le système étudié est complexe, plus il devient coûteux en termes de temps de calculs pour implémenter la dimension de Hausdorff. Si on décide en revanche d'assouplir la contrainte qui cherche à minimiser l'ensemble des parties qui recouvre la fractale \mathcal{A} , et que l'on se contente de recouvrir cette dernière par une grille de taille fixe (Figure 2.2), on obtient alors une limite supérieure de la dimension de Hausdorff que l'on appelle « dimension de capacité »,

ou « dimension box-counting » [107]. Il est important de préciser que, pour la plupart des fractales, la dimension de capacité et la dimension de Hausdorff sont égales. Si on reprend l'équation (2.2) et que l'on considère r comme étant la taille de la grille, alors (2.2) devient :

$$\Gamma(\mathcal{A}, D, r) = \sum_i \delta_i^D = \sum_i r^D = n(r) r^D \quad (2.4),$$

où $n(r)$ correspond au nombre non vide de carrés de la grille. La dimension de comptage du nombre de carrés non vide de la grille est la valeur de D correspondant à la transition entre $\Gamma \rightarrow \infty$ et $\Gamma \rightarrow 0$. Ici, lorsque $\Gamma(\mathcal{A}, D, r) \sim 1$, cela implique que ,

$$n(r) \sim r^{-D_H} \quad (2.5),$$

et de manière plus générale :

$$D_H = \lim_{r \rightarrow 0} \frac{\log(n(r))}{\log(1/r)} \quad (2.6).$$

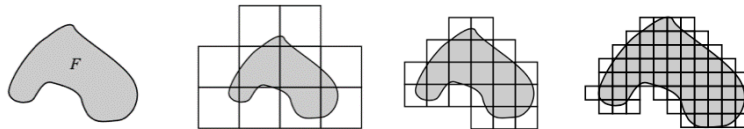


Figure 2. 2 Exemple d'une mise à l'échelle d'un système avec un système de grille pour déterminer la dimension avec la méthode de « box-counting ». (google.com/images.com)

2.2.3 Dimension de Corrélation

Cette dimension donne une définition alternative de la dimension fractale. Elle fut introduite en 1983 par Grassberger et Procaccia [56] et propose une méthode de mesure de la complexité d'un système complexe de sorte à pouvoir distinguer des systèmes dits parfaitement déterministes et chaotiques des systèmes parfaitement aléatoires. Considérons une série $x(n)$, avec $n = 1, \dots, N$ et N le nombre total de points dans la série. On peut alors définir un vecteur de dimension m que l'on nomme $y_m(i)$ qui est généré par la série suivante :

$$y_i^m = [x_i, x_{i+\tau}, x_{i+2\tau}, \dots, x_{i+(m-1)\tau}]^T \quad (2.7),$$

où τ correspond à un temps de retard [45]. On définit également le nombre de vecteurs $y_m(i)$ par $N_m = N - \tau(m - 1)$ pour chaque dimension m . La distance entre chaque paire de vecteurs $\{y_i^m, y_j^m\}$ est définie par :

$$d_{i,j}^m = d(y_i^m, y_j^m) \quad (2.8)$$

et peut être calculée comme étant la norme de la différence de ces deux vecteurs à savoir $\Delta y_{i,j}^m = y_i^m - y_j^m$. La somme de corrélation, qui représente la probabilité que la distance de paire de vecteurs $d_{i,j}^m$ soit inférieure à un certain seuil r , est calculée comme suit :

$$C_m(r) = \frac{1}{N_m(N_m-1)} \sum_{i,j=1}^{N_m} H(r - d_{i,j}^m) \quad (2.9)$$

$$C_m(r) = \frac{1}{N_m(N_m-1)} \sum_{i=1}^{N_m} C_i^m(r) \quad (2.10),$$

avec $H(\cdot)$ la fonction de Heaviside définie par : $H(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases}$ et $C_i^m(r) = \sum_{j=1}^{N_m} H(r - d_{i,j}^m)$.

Grassberger et Procaccia ajoutent que, pour des systèmes purement déterministes, $C_m(r)$ décroît de manière monotone vers 0 lorsque r tend vers 0 et N_m tend vers l'infini, alors $C_m(r)$ peut être très bien approximé par la simplification suivante :

$$\lim_{N_m \rightarrow \infty} \lim_{r \rightarrow 0} C_m(r) = r^{D_2^m}$$

avec D_2^m la dimension de corrélation pour un vecteur de taille m . Ainsi, on peut définir de manière plus générale la dimension de corrélation par la relation suivante :

$$D_2^m = \lim_{r \rightarrow 0} \frac{\log C_m(r)}{\log(r)} \quad (2.11),$$

ce qui implique que, lorsque l'on trace $\log C_m(r)$ en ordonné contre $\log(r)$ en abscisse, on devrait obtenir un gradient constant D_2 lorsque $r \rightarrow 0$. Et plus on augmente la dimension m du vecteur, à savoir $N_m \rightarrow \infty$, plus les valeurs de D_2^m tendent à saturer vers une valeur finie et ainsi à converger vers la valeur de D_2 qui constitue l'estimation finale de la dimension de corrélation telle que :

$$D_2 = \lim_{N_m \rightarrow \infty} \lim_{r \rightarrow 0} \frac{\log C_m(r)}{\log(r)} \quad (2.12).$$

Une limite importante concernant la dimension de corrélation est également le coût de calcul nécessaire à la convergence vers la valeur finale de D_2 du fait de l'estimation de la somme des corrélations des séquences qui peut être plus ou moins fastidieuse selon le choix de la dimension m et la taille du vecteur y_i . C'est pourquoi, afin de contourner ce problème de coût de calcul lié à l'approche traditionnelle de mesure géométrique, nous proposons d'introduire une approche statistique plus efficace et plus simple d'implémentation communément appelée la Dynamique Symbolique.

2.3 Dynamique Symbolique et Théorie de l'information de Shannon

2.3.1 Dimension de Shannon :

Parmi les précédentes méthodes d'estimations de la dimension fractale des systèmes complexes, nous montrons que leur approche respective reste similaire. En effet, l'idée est de considérer des quantificateurs de dimensions comme étant la masse d'une sphère de diamètre r (ou une grille de carrés de longueur r pour la dimension de capacité) qui couvre une partie d'une figure fractale, selon des facteurs d'échelles invariants de sorte que la masse nommée $p(r)$ soit approximée par une loi de puissance pour des valeurs de r proches de 0, telles que $p(r) \sim r^D$, avec D étant la dimension associée à cette masse. C'est donc à chaque fois une mise à l'échelle de la masse avec la taille afin de s'ajuster le plus fidèlement possible à la forme géométrique de la fractale considérée.

Une alternative à cette méthodologie est de considérer la dimension fractale d'un ensemble en termes de nombres réels nécessaires pour spécifier un point sur cet

ensemble. Par exemple, la position d'un point sur une ligne peut être étiquetée par un seul nombre réel ; la position sur un plan par deux coordonnées cartésiennes ; la position dans l'espace ordinaire (tridimensionnel) par trois coordonnées. Ici, on considère la dimension en termes de nombres de degrés de liberté que possède un ensemble. Pour les ensembles plus complexes que de simples lignes, simples surfaces ou volumes, cette définition doit être étendue. C'est en ces termes qu'intervient la dimension d'information dite de Shannon.

Une manière d'élargir cette définition est de déterminer non pas combien de nombres réels mais combien de bits d'informations sont nécessaires pour spécifier un point d'un ensemble avec une précision donnée. Sur un segment de longueur unitaire, k bits sont nécessaires pour spécifier la position d'un point avec une précision $r = 2^{-k}$. Pour un carré de côté de longueur unitaire, $2k$ bits sont nécessaires pour obtenir la même précision (k bits pour chacune des deux coordonnées spécifiées). Et, de la même manière, en ce qui concerne un cube d'arrête de longueur unitaire, $3k$ bits sont nécessaires.

De façon plus générale Shannon montre qu'il faut $I(r) = -d \log_2(r)$ bits d'informations pour pouvoir spécifier la position d'un hyper-cube de dimension d unitaire avec une précision de r . On note que $I(r)$ est l'information (en bits) nécessaire pour spécifier un point sur l'ensemble avec une précision r .

Shannon considère $I(r)$ comme le degré d'incertitude que l'on a sur un système et postule que $2^{-I(r)}$ est le nombre total d'états disponibles de ce même système. On peut interpréter $2^{-I(r)}$ comme la masse moyenne de chaque état de l'ensemble. Ainsi, on retrouve également la notion de masse mais en termes de probabilité, ce qui nous permet de faire le lien avec les autres dimensions évoquées ci-dessus, en interprétant la dimension de l'information comme étant aussi une mise à l'échelle du volume avec la taille.

Si on décide de partitionner l'ensemble fractal considéré en boîtes \mathcal{B}_i avec $i = 1, \dots, N$ et N le nombre total de boîtes de taille r , alors, spécifier la position d'un point

avec une précision de taille r , nécessite que l'on spécifie dans quelle boîte se trouve le point. L'information moyenne nécessaire pour spécifier cette boîte est donnée par la formule de Shannon :

$$I(r) = -\sum_i P_i \log_2 P_i \quad (2.13),$$

où P_i mesure la probabilité de la i -ème boîte telle que :

$$P_i = \mu(\mathcal{B}_i) / \mu(\mathcal{A}) \quad (2.14),$$

avec \mathcal{A} étant la fractale. La relation (2.14) conduit directement à une expression de la dimension de l'information de Shannon de l'ensemble fractal :

$$D_I = \lim_{r \rightarrow 0} \frac{-I(r)}{\log_2(r)} = \lim_{r \rightarrow 0} \frac{\sum_i P_i \log_2 P_i}{\log_2(r)} \quad (2.15).$$

Ainsi, la dimension d'information, ou dimension de Shannon correspond au degré d'incertitude que l'on a concernant une partie d'un ensemble local de la fractale considérée. Cette information est intrinsèquement liée à la notion d'Entropie de Shannon et ceci fera l'objet de la section 2.3.3.

2.3.2 *Dynamique Symbolique*

Nous avons montré précédemment que l'une des conditions au calcul de la Dimension de Shannon est de discrétiser l'espace des phases du système en nombre de bits d'informations. Quelle méthode adopter pour cette discrétisation de l'espace ?

La dynamique symbolique étudie un système complexe basé sur les séquences de symboles obtenues pour une partition appropriée de l'espace d'états. L'idée fondamentale de la dynamique symbolique est de diviser l'espace des phases d'un système en un nombre fini de régions (partitions) et d'étiqueter chaque région par un symbole alphabétique ou numérique [36, 40, 59]. Si on imagine un point suivant une certaine trajectoire dans l'espace et que l'on décide de séquencer l'espace en un nombre de partitions fini de morceaux de sorte que chaque partition soit définie par un symbole

particulier, on obtient alors ce qu'on appelle une « trajectoire symbolique » en écrivant la séquence de symboles correspondant aux éléments de partition successifs visités par le point sur son orbite (i.e. le long de sa trajectoire). Selon la nature du système dynamique étudié et également selon le choix du type de partition choisi, la trajectoire symbolique peut déterminer le comportement de l'orbite, trouver une description simple de l'ensemble de toutes les trajectoires symboliques, et peut nous aider à comprendre la nature du dynamisme du système telle que son niveau de complexité, d'ordre et de désordre.

La première étape consiste à construire la séquence de symboles à partir du système dynamique considéré. La séquence de symboles ne peut être générée que par une partition appropriée de l'espace des phases. Cette partition est proche de la méthode de Hausdorff explicitée ci-dessus. L'idée est de recouvrir l'ensemble des phases d'un système X par un nombre total R de cellules ou de boîtes de longueur ξ . Chaque cellule n 'a aucun point de l'espace en commun et est définie par C_i avec l'index $i = 1, \dots, R$.

Ainsi nous avons :

$$\bigcup_{i=1}^R C_i = X \quad (2.16).$$

Dans le cadre d'une série temporelle, comme une série financière de rendements boursiers, on définit la première valeur r_0 de la série comme appartenant à la cellule i_0 , la deuxième valeur r_1 appartenant à la cellule i_1 et ainsi de suite. Au fur et à mesure que l'état du système dynamique évolue dans le temps, la trajectoire des rendements boursiers se transforme en une suite de séquences de symboles correspondant à des partitions, des zones de l'espace des phases du type : $\{i_0, i_1, i_2, \dots, i_R\}$. On peut ensuite encoder chaque cellule en fonction de la région de l'espace des phases à laquelle elle appartient. Pour une série financière du type rendements d'un actif boursier, on définit deux zones : une zone $I_0 = [0; +\infty[$ où les rendements sont positifs ou nuls, et une zone $I_1 =]-\infty; 0[$ où les rendements sont négatifs. Lorsque les itérations du système entrent dans le premier intervalle I_0 , le symbole 1 est émis, et lorsqu'ils entrent dans l'intervalle I_1 le symbole 0 est émis.

Ainsi, le dynamisme de la suite logistique peut donc être représenté par une séquence de 1 et de 0, indiquant sa position approximative dans une des partitions de son espace des phases pour chaque itération. On retrouve ici, la notion de bits d'information énoncée dans la section 2.2.1. Selon la taille de partition des séries de séquences, des formes, qui se répètent de manière identique au cours du temps avec un niveau de probabilité, peuvent émerger. Et c'est dans ce contexte que les recherches récentes s'intéressent à la symbolique dynamique comme un moyen d'extraire de l'information sur des systèmes plus complexes grâce à l'analyse statistique des structures émergentes des séries de symboles. L'une des études les plus célèbres concerne celle de Steve Smale en 1960 avec son article sur le fer à cheval [104, 105] ; elle montre que les systèmes complexes du type chaotique avec sensibilité aux conditions initiales sont structurellement stables et que la complexité est naturellement associée à une certaine stabilité relative. Ainsi, l'exemple du fer à cheval peut être étudié au travers des partitions de son espace de phases. Au travers de sa dynamique symbolique, l'étude démontre qu'il est possible que ce dernier fasse émerger des structures qui se répètent au cours du temps et qui, statistiquement parlant, peuvent être mesurées.

Ainsi, la dynamique symbolique est un moyen d'encoder n'importe quel type de signal en une série de séquences de symboles, desquels il est possible d'extraire un niveau d'information statistique. C'est d'ailleurs avec l'émergence de la théorie de l'information et de la théorie mathématique de la communication développée par C.E. Shannon que la dynamique symbolique s'est vu octroyer un cadre mathématique sur lequel s'appuyer.

2.3.3 Information et Entropie de Shannon :

La théorie de l'information est une représentation mathématique des conditions et des paramètres affectant la transmission et le traitement de l'information. La véritable naissance de la théorie moderne de l'information remonte à la publication en 1948 de Claude Shannon [101]. Une étape importante dans le travail de Shannon a été de réaliser que, pour établir une théorie, les signaux de communication doivent être traités indépendamment de la signification des messages qu'ils transmettent. Cette

vision contraste fortement avec la conception de l'information, dans laquelle le sens a un rôle essentiel. Shannon montre que la quantité de connaissances véhiculée par un signal n'est pas directement liée à la taille du message.

Ainsi, la théorie de l'information se focalise sur les signaux et sur la manière dont ils peuvent être transmis avec précision et efficacité ; les questions de sens sont évitées autant que possible. L'idée est de transmettre un message au travers d'un encodage le plus simple possible tout en préservant l'information initiale dans son intégralité. C'est dans ce but que Shannon développe les prémices de l'analyse Symbolique, à savoir convertir un signal continu en signal discret tout en préservant ses caractéristiques intrinsèques. L'utilisation de la Dynamique Symbolique en théorie de l'Information est fondamentale notamment pour analyser le niveau d'information contenu dans un message ou signal, quantifier son niveau de prédictibilité, son niveau de complexité et encoder ce dernier de manière efficace et sans perte d'informations. Lorsque que l'on souhaite encoder un message pour le transmettre de X vers Y Shannon propose de minimiser le nombre de bits nécessaires afin de faciliter le temps de calculs lors du décodage du message. Il propose de mesurer l'efficacité d'un message en termes de longueur moyenne des messages. Pour minimiser le nombre de bits nécessaires à la transmission du message, il faut au préalable définir la fréquence d'occurrence de mots dans un message.

Autrement dit, Shannon calcule le nombre moyen de bits utilisés pour envoyer des messages de X vers Y et cherche à minimiser cette valeur moyenne.

On considère un message comme un ensemble d'événements possibles dont les probabilités d'occurrence sont p_1, p_2, \dots, p_n avec n le nombre d'événements. Ces probabilités sont connues et représentent la seule information que nous avons sur l'événement ou le message qui sera émis par la source d'information. Dans le cas de l'exemple précédent sur les séries financières, l'information représente la probabilité qu'une certaine partition de l'espace des phases du système se produise. Shannon propose une mesure de « choix » dans l'attribution du nombre de bits pour chaque événement permettant de minimiser la taille d'encodage moyenne des messages ; il

appelle cette mesure « entropie » représentée par $H(p_1, p_2, \dots, p_n)$ et définie par les propriétés suivantes :

- H est continu pour tout p_i
- Si tous les p_i sont égaux tels que $p_i = \frac{1}{n}$, alors H devrait être une fonction croissante monotone de n . Avec des événements équiprobables, il y a par conséquent plus de choix possibles, plus d'incertitudes lorsque le nombre de possibles événements augmente.
- $H = -K \sum_{i=1}^n p_i \log p_i$ avec K une constante positive et équivaut simplement au choix d'unité de mesure.

Il existe donc un ensemble d'échantillons N d'événements auxquels on attribue des probabilités d'occurrence p_i . En discrétisant cet ensemble en R sous-ensembles disjoints, chaque sous-ensemble i est considéré comme un événement composé de N_i événements élémentaires tels que définis par Shannon:

$$N = \sum_{i=1}^R N_i \quad (2.17).$$

La probabilité de l'événement composé i , à savoir la probabilité de recevoir comme message un événement élémentaire de probabilité p_i est donnée par :

$$p_i = N_i/N \quad (2.18).$$

Comme évoqué précédemment, pour sélectionner un événement élémentaire dans l'ensemble d'échantillons, il faut toujours un nombre minimal de bits de $\log_2 N$, avec N étant le nombre d'événements différents de l'échantillon exprimé en bits. Le nombre minimal de bits requis pour sélectionner un événement élémentaire de probabilité p_i du sous-ensemble i dans lequel il se situe est $\log_2 N_i$. Si on désigne le nombre de bits nécessaires (la série de bits nécessaires pour encoder un message) pour sélectionner le sous-ensemble i par s_i , tel que ce dernier vérifie la relation suivante :

$$s_i = \log_2 N - \log_2 N_i \quad (2.19),$$

en substituant l'équation (2.18) dans (2.19) on obtient :

$$s_i = -\log_2 p_i \quad (2.20).$$

Puisque s_i mesure un manque d'information sur l'occurrence de l'événement suivant, l'opposé de s_i mesure une certaine connaissance. En combinant cette information avec la manière dont nous calculons la taille d'encodage minimale en bits (à savoir en la pondérant par le niveau de probabilité associé à l'information émise par cet événement s_i), on obtient la relation de l'entropie de Shannon comme suit :

$$H(p_i) = \sum_i p_i * s_i = -\sum_i p_i * \log_2 p_i \quad (2.21).$$

En d'autres termes, l'Entropie de Shannon correspond au manque d'informations sur l'occurrence d'un événement futur, à savoir $-I(p_i)$ caractérisé par une probabilité p_i .

Puisque l'entropie de Shannon correspond à l'opposé de l'information, on peut assimiler l'entropie à la notion de désordre, d'incertitude, d'imprévisibilité, mais aussi à une autre forme de mesure de l'information. Si l'entropie est élevée, à savoir que la taille de l'encodage est grande en moyenne, cela signifie que nous avons de nombreux types de messages avec de petites probabilités. Par conséquent, chaque fois qu'un nouveau message arrive, on peut s'attendre à un type de message différent du précédent. On peut comprendre cela comme une incertitude, une imprévisibilité. Lorsqu'un type de message avec une probabilité beaucoup plus faible que d'autres types de message se produit, ce dernier est considéré comme un phénomène de surprise, car, en moyenne, on s'attend à voir des messages plus fréquents.

En outre, un type de message rare a plus d'informations au sens de Shannon que les types de message plus fréquents, car ils éliminent de nombreuses autres probabilités et fournissent plus d'informations spécifiques.

Ainsi, si l'entropie est élevée, la taille de codage moyenne est significative, ce qui signifie que chaque message a tendance à avoir plus d'informations spécifiques au sens de Shannon. C'est pourquoi on associe une entropie élevée au concept de désordre,

d'incertitude, phénomène de surprise, à l'imprévisibilité, ou encore à la quantité d'informations.

En revanche, une faible entropie signifie que, la plupart du temps, nous recevons les informations les plus prévisibles, signifiant moins de désordre, moins d'incertitude, moins de surprise, plus de prévisibilité et moins d'informations spécifiques au sens de Shannon. Nous proposons maintenant d'introduire une définition Générale de la mesure de l'information qui est une généralisation de l'entropie de Shannon à savoir l'entropie de Rényi.

2.4 Dimension et Entropie Généralisée

2.4.1 Dimension de Rényi

Parmi les différentes définitions de dimensions fractales évoquées ci-dessus, notamment celles qui concernent la dimension de capacité et de corrélation, le calcul de la dimension résulte d'un comptage de cases ; la décision de compter ou pas la case dépend de la présence ou pas de points de l'ensemble fractal dans ces cases. Aucune disposition n'est prise pour pondérer le nombre de cases en fonction du nombre de points présents dans une case. En d'autres termes, la structure géométrique de l'ensemble fractal est analysée, mais la mesure sous-jacente est ignorée, ce qui est une autre limite à l'utilisation de ces définitions, outre le fait qu'elles soient coûteuses en termes de calculs. De plus, ces dimensions sont utiles uniquement pour les ensembles dit mono-fractals (self-similaires).

La dimension généralisée [58, 62], aussi appelée « dimension de Rényi » [96] ou encore « dimension d'information de Rényi » est ainsi nommée car elle permet de généraliser toutes les définitions précédentes, notamment la dimension d'information de Shannon. Elle permet de prendre en compte le nombre de points dans l'ensemble des cases qui recouvrent l'ensemble fractal.

Considérons comme précédemment \mathcal{B}_i comme étant la i ème case qui recouvre l'ensemble fractal \mathcal{A} , et définissons $P_i = \mu(\mathcal{B}_i)/\mu(\mathcal{A})$ la mesure normalisée de cette case. De manière équivalente, c'est la probabilité qu'un point choisi aléatoirement sur

l'ensemble fractal \mathcal{A} soit dans la i ème case \mathcal{B}_i ; cette probabilité est généralement estimée en comptant le nombre de points qui se trouvent dans i ème case et en divisant par le nombre total de points.

Ainsi, en divisant l'espace des phases de l'ensemble \mathcal{A} , on obtient un espace des phases de dimension d . Nous excluons les cases qui ont une probabilité égale à zéro et étiquetons celles avec une probabilité non nulle par $i = 1, 2, \dots, R$. Le nombre de carrés, avec une probabilité non nulle r , se distingue du nombre total de cases par la relation : $R \sim r^{-d}$. Outre le fait de garder le nombre de carrés total R fini, nous nous intéressons au comportement limite de la taille des carrés lorsque $r \rightarrow 0$, ce qui implique que $R \sim r^{-d} \rightarrow \infty$. Quand r tend vers zéro, l'information de Rényi se définit par la relation :

$$I_q(p) = \frac{1}{q-1} \ln \sum_{i=1}^r (p_i)^q \quad (2.22),$$

avec q un nombre arbitraire, r est le nombre de cases non vides qui recouvre l'ensemble fractal pour un état i . Autrement dit, i parcourt toutes les probabilités non nulles et tous les événements avec une probabilité nulle sont exclus.

De même, on introduit la dimension de Rényi [96] D_q comme le changement d'information de Rényi pour $r \rightarrow 0$ tel que :

$$D_q = \frac{1}{q-1} \lim_{r \rightarrow 0} \frac{\log \sum_i p_i^q}{\log(r)} \quad (2.23).$$

La dimension de Rényi est une fonction décroissante de q et restera finie tant que $r \rightarrow 0$. En fonction de la valeur de q , la dimension de Rényi révèle des caractéristiques différentes de l'ensemble fractal. Par exemple, pour $q = 0$, l'information négative de Rényi devient égale au logarithme du nombre total de cases non nulles, à savoir le nombre R . Cela correspond au nombre minimum de cases pour couvrir la fractale dans son ensemble, d'où, pour $q = 0$, on obtient :

$$D_0 = - \lim_{r \rightarrow 0} \frac{\ln R}{\ln(r)} = \lim_{r \rightarrow 0} \frac{\ln R}{\ln(1/r)} \quad (2.24),$$

et D_0 correspond à dimension fractale de capacité (« box-counting ») évoquée dans la section 2.2.2.

Pour $q = 1$, l'information de Rényi définie par l'équation (2.22) devient l'information de Shannon. Ainsi on peut définir la dimension de Rényi pour $q = 1$ comme suit :

$$D_1 = \lim_{r \rightarrow 0} \frac{1}{\ln(r)} \ln \sum_{i=1}^R p_i \ln(p_i) \quad (2.25),$$

qui décrit comment l'information de Shannon varie en fonction de la taille des carrés qui recouvrent l'ensemble fractal.

Et enfin pour $q = 2$, on retrouve la dimension de corrélation telle que :

$$D_2 = - \frac{\ln \sum_{i=1}^R p_i^2}{\ln(r)} \quad (2.26).$$

Ainsi, les dimensions de Hausdorff, de capacité, de corrélation et de Shannon sont des cas particuliers pour des valeurs de q de la dimension de Rényi.

On note que si $D_0 = D_1 = D_2$, alors le système considéré se caractérise par une structure fractale simple, et on appelle cet ensemble une mono-fractale ou uni-fractale.

En revanche, si $D_0 > D_1 > D_2$, le système se caractérise par une structure fractale dite complexe et l'objet est appelé multi-fractal [91].

De manière générale, q détermine le poids ou l'importance attribués à la mesure de la probabilité p . Quand $q > 0$, D_q est principalement influencé par les cases à forte probabilité, et lorsque $q < 0$, les cases avec les probabilités les plus faibles influencent le plus la dimension de Rényi. Ainsi, si le spectre de dimension de Rényi (aussi appelé spectre multi-fractal) décroît lorsque q augmente, alors le système peut être considéré comme étant une structure multi-fractale ; dans la négative, si le spectre de dimension reste constant pour toutes les valeurs de q , alors le système sera considéré comme une structure fractale simple mono-fractale.

Nous reviendrons sur cette notion de spectre multi-fractal dans la section 2.4.2 lorsqu'on abordera le spectre de Rényi aussi appelé spectre de singularité.

Néanmoins, l'idée fondamentale reste de pouvoir caractériser de manière plus précise les fractales avec une mesure non uniforme. Au lieu d'avoir une seule mesure de la dimension pour un système, on peut calculer le spectre complet des dimensions généralisé à partir de $D_{-\infty}$ jusqu'à $D_{+\infty}$. De cette manière on peut généraliser le formalisme de la mesure granulaire (« coarsed-grained ») introduit dans la section 2.2.1 concernant la dimension de Hausdorff. En généralisant l'équation (2.2) on définit :

$$\Gamma_q(\mathcal{A}, D, r) = \sum_i \mu_i^q \delta_i^{(1-q)D} \quad (2.27),$$

avec $\mu_i = \mu(\mathcal{B}_i)$ la mesure du ième élément du revêtement de la fractale.

Comme défini dans la section 2.2.1, lorsque $r \rightarrow 0$, nous obtenons $\Gamma_q(\mathcal{A}, D) \rightarrow 0$ pour $(1 - q)D < (1 - q)D_q$ et $\Gamma_q(\mathcal{A}, D) \rightarrow \infty$ pour $(1 - q)D > (1 - q)D_q$.

On note que la transition pour passer de $\Gamma \rightarrow 0$ vers $\Gamma \rightarrow \infty$ définit la dimension généralisée :

$$D_q = \frac{1}{1-q} \inf\{ (1 - q)D \lim_{r \rightarrow 0} \Gamma_q(\mathcal{A}, D, r) = 0 \} \quad (2.28).$$

L'avantage des équations (2.27) et (2.28) est qu'elles fournissent une définition de la dimension généralisée sans nécessiter des cases de taille fixe.

2.4.2 Spectre de singularité

Le concept de singularité de force α fut développé pour la première fois grâce à l'agrégation des travaux de recherche Hasley, Meankin et Procaccia, notamment au travers d'un article publié en 1985 [61]. Ces derniers introduisent un changement de variables qui permet d'avoir une nouvelle interprétation de la dimension généralisée de Rényi, au travers de la densité de singularité α qu'ils expriment en termes de fonction $f(\alpha)$.

Afin de déterminer la fonction de densité $f(\alpha)$ (spectre de singularité α) pour une mesure donnée, Hasley [61] propose de la relier aux propriétés observables de la mesure. On relie $f(\alpha)$ à un ensemble de dimensions que nous avons introduites dans la section 2.4.1, à savoir la dimension généralisée de Rényi D_q définie sous sa forme générale par l'équation (2.23). En considérant la dimension de Rényi, la méthode d'Hasley consiste à faire un changement de variables au travers d'une transformation de Legendre partant des variables $(q, D_q(q - 1))$ vers les nouvelles variables (α, f) .

Dans un premier temps, posons quelques définitions afin de pouvoir introduire le concept de spectre de singularité [100]. Considérons un sous-ensemble, un maillage constitué de plusieurs cases d'unité r couvrant un domaine $D \subset \mathbb{R}^n$ sur lequel est défini une fractale caractérisée par une mesure multi-fractale finie μ . Cette mesure multi-fractale μ se caractérise par une distribution telle qu'autour de tout point $x \in D$, la mesure d'une case du maillage de longueur unitaire r autour de x est proportionnelle à la loi de puissance r^α , avec α correspondant à l'exposant de mise à l'échelle et peut prendre des valeurs quelconques, à condition que r soit suffisamment petit, de telle sorte que les ensembles formés par les cases du maillage soient définis par le même exposant d'échelle α .

De manière plus générale, la mesure multi-fractale μ pour des valeurs de r suffisamment petites se note :

$$\mu_r(x) \sim r^{\alpha_x} \quad (2.29)$$

pour un facteur d'échelle α_x pour tout $x \in D$, où $\mu_r(x)$ est la mesure d'une case de longueur unitaire r autour d'un point x .

On peut également interpréter la mesure multi-fractale $\mu_r(x)$ d'une manière probabiliste. En effet, si on appelle N le nombre de cases du maillages nécessaires à recouvrir le sous-ensemble D , alors la probabilité d'observer une case quelconque est définie par la relation suivante :

$$p_i := \frac{N_i}{N} = \int_{i\text{ème cases}} d\mu(x) \quad (2.30),$$

avec p_i la probabilité qu'une partie de l'ensemble fractal soit mesurée dans la i ème case, et μ une mesure multi-fractale finie sur l'ensemble D . Ainsi, avec les équations (2.29) et (2.30) on peut interpréter la règle d'échelonnage α par les deux relations suivantes [62, 63] :

$$p_i \sim r^{\alpha_i} \quad (2.31),$$

et si on considère que α peut prendre des valeurs définies par l'intervalle $[\alpha_i, \alpha_i + d\alpha_i]$ correspondant à des régions différentes de mesures de $\mu_r(x)$; en particulier, si le système est divisé par des cases de longueur r , alors le nombre de fois où α sera compris entre des valeurs de l'intervalle α_i et $\alpha_i + d\alpha_i$ sera défini par $N(\alpha_i)$ tel que :

$$N(\alpha_i) \sim \rho(\alpha_i) d\alpha_i r^{-f(\alpha_i)} \quad (2.32),$$

avec $f(\alpha_i)$ une fonction continue qui reflète les dimensions différentes des sous-ensembles définis par les singularités α_i et ρ et correspond à la fonction de densité utilisée pour prendre en compte la dimension D . Une fois que (2.31) et (2.32) sont définis, nous sommes capables de calculer efficacement le spectre $f(\alpha)$ par la méthode probabiliste [62, 63, 47, 8]. En élevant p_i à la puissance q pour des valeurs quelconques de q , comme défini par la dimension de Rényi, suivant l'équation (2.31) on obtient que :

$$Z(q) := \sum_i p_i^q \sim \sum_i r^{\alpha_i q} \sim \int_{\alpha} N(\alpha) r^{\alpha q} \sim \int \rho(\alpha) r^{\alpha q - f(\alpha)} d\alpha \quad (2.33),$$

et pour des valeurs suffisamment petites de r , la valeur de $Z(q)$ est principalement déterminée par la singularité α , et est équivalente à la valeur minimale de :

$$\tau(q) = \min_{\alpha} [\alpha q - f(\alpha)] \quad (2.34).$$

Appelons $\alpha(q)$ cette valeur de α et, par transformation de Legendre, nous obtenons de manière triviale que :

$$\alpha(q) = \frac{d\tau(q)}{dq} \quad (2.35),$$

et :

$$f(\alpha(q)) = \alpha(q)q - \tau(q) \quad (2.36)$$

que l'on peut aussi écrire sous la forme :

$$f(\alpha) = \min_q [\alpha q - \tau(q)] \quad (2.37).$$

Ainsi, en faisant varier $\tau(q)$ pour $q \in [-\infty; +\infty]$, on obtient le spectre de singularité $f(\alpha(q))$ dans son intégralité. En revanche, il n'est pas possible en pratique d'utiliser des valeurs infinies de q , ce qui n'est pas non plus souhaitable puisque, pour des valeurs extrêmes de q , nous perdons en fiabilité. Afin de sélectionner un intervalle de q approprié, il faut définir un seuil limite pour l'erreur générée par l'ajustement linéaire et rejeter toutes les valeurs de q pour lesquelles le seuil est dépassé. Il est aussi nécessaire de sélectionner q de sorte que le spectre de singularité $f(\alpha) > 0$ et que la dimension généralisée de Rényi D_q reste inférieure à la dimension du d de l'ensemble fractal défini ci-dessus par le sous-ensemble D ; cela est particulier vrai pour des valeurs négatives de q . Ainsi, on souhaite restreindre les valeurs de q de sorte que $D_q < d$; dans le cas contraire, D_q perdrait toute significativité.

En pratique, $\tau(q)$ est déterminé directement comme étant la pente du graphique log-log de $\sum_i \mu_i^q(r)$ contre r obtenu pour des tailles de cases de maillage de longueur unitaire r différentes, avec $\mu_i^q(r)$ correspondant aux nombre total de cases i de longueur r suffisamment petites pour recouvrir le sous-ensemble D . Puisque cette pente est indépendante à la normalisation de la mesure de μ , μ ne nécessite pas d'être pondéré par une mesure de probabilité. De manière plus explicite, lorsque la longueur r des cases du maillage tend vers zéro alors $\tau(q)$ correspond à la somme normalisée logarithmique de l'ensemble des carrés du maillage telle que :

$$\tau(q) = \lim_{r \rightarrow 0} \frac{\log(\sum_i \mu_i^q(r))}{\log(r)} \quad (2.38).$$

Finalemment, en considérant la dimension de Rényi définie par l'équation (2.23), on peut donner une autre expression à cette dernière en substituant (2.38) dans (2.23) :

$$D_q = \frac{\tau(q)}{1-q} \quad (2.39),$$

et pour $q \neq 1$, à partir de l'équation (2.34) on obtient :

$$\left\{ \begin{array}{l} D_q = \frac{f(\alpha) - q\alpha}{1-q} \quad (2.40) \\ f(\alpha) = D_q(1-q) + q\alpha \quad (2.41) \end{array} \right.$$

Cependant, déterminer $\tau(q)$ à la fois au travers d'un ajustement linéaire et d'une application numérique dite de Legendre a un coût non négligeable sur la précision des résultats. Chhabra et Sreenivasan [30] proposent qu'une manière efficace de contourner ce problème est de moyennner plusieurs échantillons, soit sur une plage d'échantillons $\{f_j(\alpha_i)\}_j$ calculé indépendamment pour différents échantillons j , soit en calculant la moyenne d'abord sur une plage de $\{N_j(\alpha_i)\}_j$ et ensuite déduire la valeur correspondante de $f(\alpha_i)$ grâce à la relation définit par équation (2.32). Les deux auteurs privilégient la deuxième solution afin de garantir une certaine fiabilité ; en revanche cette solution risque de créer artificiellement des valeurs négatives dans le spectre de singularité si $N(\alpha_i)$ prend des valeurs inférieures à 1 pour certaines valeurs de α_i comme résultats au processus de calcul de la moyenne. Pour remédier à ce problème, une méthode de multiplication est présentée par Chhabra et Sreenivasan [30]. L'idée est de considérer que l'auto-similitude de la mesure fractale implique l'existence d'une distribution d'échelle sous-jacente telle que les α_i , pour un niveau de mise à l'échelle donné r_k , soit simplement le résultat d'un effet multiplicateur de k compositions de mise à l'échelle. S'il n'y a pas de corrélation dans le processus probabiliste sous-jacent de la résolution r_{k-1} à la résolution r_k pour une valeur quelconque de k , alors on peut en déduire que les multiplicateurs sont les α . En

particulier, si tous les niveaux de résolution ne sont pas corrélés, on peut choisir $k = 1$; si non, il faut choisir la plus petite valeur de k possible pour laquelle un niveau de résolution est non corrélé au précédent. Si on appelle r_0 la résolution minimale et r_k la résolution choisie, alors on peut définir :

$$r := \frac{r_k}{r_0} \quad (2.42),$$

et pour chaque échantillon j et case i , on définit : $M_{ij} := \mu_{ij}(r_0)/\mu_{ij}(r_k)$. Ainsi, les fonctions $\tau(q)$ et $\alpha(q)$ sont données par [30] :

$$\frac{1}{N(r)} \tau(q) + d \sim - \frac{\log(1/N(r) \sum_{i=1}^{N(r)} M_{ij}^q)}{\log(r)} \quad (2.43)$$

et,

$$\frac{1}{N(r)} \alpha(q) + d \sim - \frac{\sum_{i,j} M_{ij}^q \log(M_{ij})}{\sum_{i,j} M_{ij}^q \log(\log(r))} \quad (2.44),$$

avec $N(r)$ le nombre non nul de valeurs de M_{ij} et d la dimension de l'ensemble D .

Q. Cheng and F.P. Agterberg [28] montrent que, seules les cases qui sont complètement délimitées par les limites de D devant être incluses dans le calcul de $\tau(q)$ et $\alpha(q)$, il faut donc s'assurer que les cases de la grille soient suffisamment petites afin de restreindre le moins possible la plage d'étude. Ainsi, ils proposent une autre manière équivalente de définir la fonction $\tau(q)$ comme suit:

$$\frac{1}{N(r)} \tau(q) + d \sim \lim_{r \rightarrow 0} \frac{\log(1/N(r) \sum_{i=1}^{N(r)} \mu_i^q(r))}{\log(r)} \quad (2.45),$$

avec $N(r)$ le nombre de cases de longueur unitaire non nulle r ; $\mu_i^q(r)$ est la mesure correspondant à i ème cases qui recouvrent l'ensemble D et d est la dimension de l'ensemble fractal D .

On peut ensuite définir nos probabilités p_i associées à chaque case de la grille de telle sorte :

$$\mu_i^q(r) = \frac{p_i^q}{\sum_j p_j^q} \quad (2.46).$$

Puis, la « transformée de Legendre » peut être directement intégrée dans le calcul de f et α grâce aux formules suivantes :

$$f(q) = \lim_{r \rightarrow 0} \frac{\sum \mu_i^q(r) \log(\mu_i^q(r))}{\log(r)} \quad (2.47),$$

$$\alpha(q) = \lim_{r \rightarrow 0} \frac{\sum \mu_i^q(r) \log p_i(r)}{\log(r)} \quad (2.48).$$

On notera que $\alpha(q)$ est la valeur moyenne de α pour une résolution donnée q . Les relations (2.47) et (2.48) sont celles qui seront utilisées dans la section 1.5 de ce même chapitre lorsque nous étudierons la complexité des séries financières associées aux indices futurs des marchés américains notamment.

Finalement, on appelle « spectre multi-fractal » ou « spectre de singularité » la courbe $f(\alpha)$ pour des valeurs de α , et cette dernière permet -comme le spectre de Rényi- de déterminer l'ensemble des dimensions fractales d'un ensemble.

2.4.3 Entropie Généralisée (Rényi & Tsallis):

L'entropie de Shannon est une mesure d'imprévisibilité d'un système physique, pouvant aussi être décrite comme étant une quantité d'informations obtenues par l'observation de systèmes complexes, et définie par l'équation 2.21. L'entropie de Shannon est une fonction décroissante d'une dispersion de variable aléatoire ; elle est maximale lorsque tous les événements sont équiprobables. L'entropie de Shannon peut être utilisée globalement pour l'ensemble des données, ou localement pour évaluer l'entropie des distributions de densité de probabilités autour de certains points. Cette notion d'entropie peut être généralisée pour fournir des informations supplémentaires sur l'importance d'événements spécifiques, par exemple des valeurs aberrantes ou des

événement rares. En comparant l'entropie des deux distributions, l'entropie de Shannon suppose implicitement un certain compromis entre les contributions des queues de distributions et la masse principale de ces distributions. Cependant, contrôler ce compromis devient primordial lorsque, par exemple, nous souhaitons distinguer un signal faible chevauchant un signal beaucoup plus fort. C'est pourquoi il est plus commun de définir une entropie plus générale qui dépend de puissances de probabilités des événements, $\sum_{i=1}^n (p_i)^q$, avec la puissance q qui joue le rôle de contrôle en donnant plus ou moins d'importance à des probabilités faibles ou élevées selon sa valeur. En effet, si q est une valeur positive grande, elle impactera davantage les événements fréquents, alors que, pour une valeur grande et négative de q , cette dernière sera plus sensible aux événements rares avec faible probabilité d'occurrence.

Par conséquent, Constantino Tsallis [116] et Alfred Rényi [96] ont tous les deux proposé des entropies généralisées qui pour $q = 1$ correspond à l'entropie de Shannon. L'entropie de Rényi est définie comme suit :

$$K_q = \frac{1}{1-q} \log(\sum_{i=1}^n p_i^q) \quad (2.49),$$

et l'entropie de Tsallis est définie comme suit :

$$S_q = \frac{1}{q-1} (1 - \sum_{i=1}^n p_i^q) \quad (2.50),$$

L'entropie de Rényi et celle de Tsallis ont des propriétés similaires à l'entropie de Shannon :

- L'entropie est additive
- Elle atteint son maximum pour $\ln(n)$ avec $p_i = 1/n$,

et contient une propriété supplémentaire q qui est un paramètre de contrôle de l'entropie de manière à la rendre plus ou moins sensible à la forme des distributions de probabilités.

Finalement, grâce à la dynamique symbolique permettant une transformation et une standardisation des données d'une série temporelle financière quelconque (au travers d'un encodage), simplifiant ainsi la mesure du niveau d'information émergent de cette même série (à l'aide notamment du spectre de singularité), nous pourrions être en mesure de reconstruire l'état causal issu de ces données financières. En effet, à partir d'une description probabiliste du comportement de la série économique, on pourra en déduire le niveau de prédictibilité et de complexité, et ceci avec un effort de calcul minime. C'est ce que nous tâcherons de démontrer dans la section suivante.

2.5 Analyse et mesure de la complexité d'une série économique au travers de la dynamique symbolique et de la théorie de l'information

Nous proposons d'examiner le caractère prévisible et complexe du comportement dynamique du Standard & Poor500 en considérant ce dernier comme un système du type « coarse-grained » en utilisant la méthode de la dynamique symbolique et la théorie de l'information au travers du concept d'entropie et d'incertitude, mis en avant ci-dessus. Nous pensons que la mesure expérimentale de l'entropie comme moyen d'examiner la complexité d'un système est plus pertinente que les tests plus courants d'universalité dans la transition vers le chaos car elle fait abstraction de tout préjugé préalable sur les causes sous-jacentes associées à la dynamique du système, peu importe qu'il soit déterministe ou stochastique. Le but de cette section est de proposer une autre approche pour mesurer et quantifier le degré de prédictibilité et de complexité d'une série chronologique économique, sans la considérer au préalable comme étant le résultat d'un système déterministe.

Pour ce faire, nous considérons la série temporelle économique comme le système complexe d'étude. Nous souhaitons déterminer la quantité d'informations qui émerge de ce dernier selon différentes échelles de temps et selon différents paramètres d'échelles spatiales.

Nous proposons une méthode similaire à celle de Shannon reposant sur la théorie de l'information qui considère le degré de complexité d'un système comme une mesure de l'incertitude. Selon Shannon, plus les partitions d'un système font émerger un certain

ordre statistique, plus le comportement de ce dernier est prévisible et moins l'occurrence des partitions futures transmettent de l'information sur la dynamique future du système.

Nous proposons d'utiliser la même approche et de l'appliquer à la série chronologique économique du Standard & Poor500, afin d'étudier le degré d'incertitude de la dynamique des variations de prix observable sur différentes échelles de temps (minutes, 4 heures et journalier) et mesurer le degré de prévisibilité de ces variations.

2.5.1 Dynamique Symbolique et Partitions : méthodologie

Dans cette section, nous étudions la série chronologique économique de l'indice Standard & Poor500, un indice boursier qui représente la valeur pondérée des 500 plus grandes entreprises aux U.S.A représentée par les Figures⁹ 2.3, 2.4 et 2.5. Plus précisément, nous considérerons les rendements logarithmiques de cet indice comme le système étudié et nous appelons $r_i = \log(P_{i+1}) - \log(P_i)$ pour $i = 0, 1, \dots, N - 1$, r_i les rendements pour chaque temps discret i et P_i le prix de l'indice pour chaque temps discret i . L'ensemble des données remonte à janvier 2016 et se termine en mars 2021. Différentes échelles de temps sont analysées, en particulier les échelles de temps petites (minutes), les échelles de temps que l'on considère moyennes (4 heures) et les échelles de temps plus grandes (journalières). Le but est de mettre en avant la dépendance de la dynamique de l'indice boursier à l'échelle temporelle. De même, le système sera partitionné en série de symboles dynamiques de tailles différentes afin de mesurer la dépendance du système à l'échelle spatiale.

En effet, si la série chronologique est un système complexe, alors elle devrait faire émerger des structures locales (des formes géométriques, des statistiques) à des échelles de temps et d'espace plus petites qui sont auto-affines ou auto-similaires aux structures présentes sur des échelles spatio-temporelles plus grandes.

⁹ Le langage informatique Python3.9 fut utilisé pour la réalisation de l'ensemble des figures.

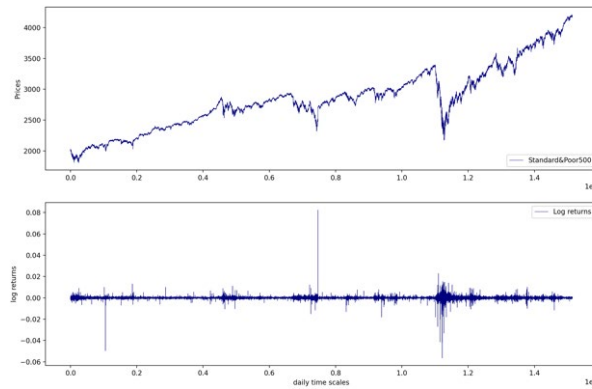


Figure 2. 3 Prix et rendements logarithmiques historiques du Standard&Poor500 entre janvier 2016 et mars 2021 sur l'échelle de temps minute.

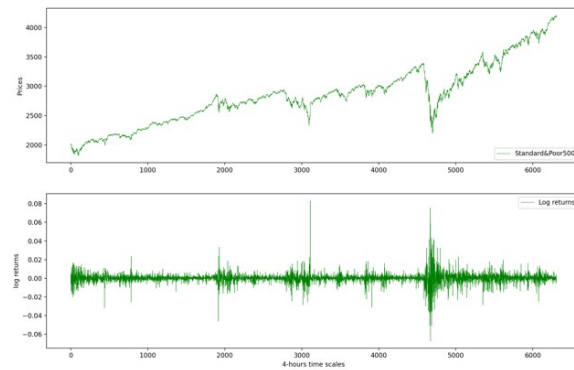


Figure 2. 4 Prix et rendements logarithmiques historiques du Standard&Poor500 entre janvier 2016 et mars 2021 sur l'échelle de temps 4-heures.

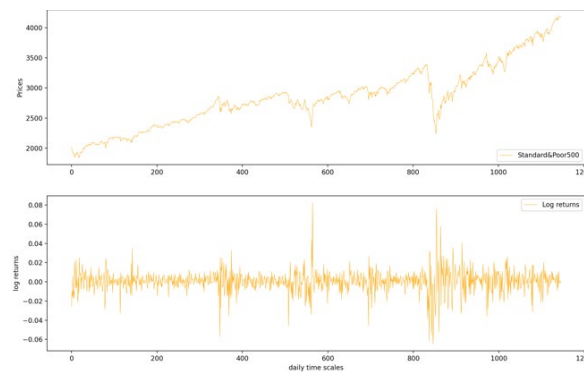


Figure 2. 5 Prix et rendements logarithmiques historiques du Standard&Poor500 entre janvier 2016 et mars 2021 sur l'échelle de temps journalière.

Tout d'abord, la méthode de la dynamique symbolique consiste à générer une séquence de symboles par une partition propre de l'espace des phases. Dans notre cas, la dynamique symbolique proviendra d'un certain nombre de choix binaires de l'espace d'états à savoir $\{0, 1\}$. En d'autres termes, si nous considérons l'intervalle d'espace des phases $(-\infty, +\infty)$, pour chaque temps discret donné i , lorsque $r_i \leq 0$, alors $r_i = 0$ et, lorsque $r_i \geq 0$, alors $r_i = 1$. Par conséquent, on peut définir un nouveau vecteur appelé x_i et on attribue à chaque x_i , pour $i = 0, 1, \dots, N - 1$ une séquence de symboles telle que : $x_i = x_0, x_1, \dots, x_{N-1}$ avec $x_i \in \{0,1\}$.

Puis, nous voulons diviser l'espace d'états de la séquence de symboles x_i en différentes partitions ou segments que nous appelons S d'égale longueur L dans lesquels on souhaite observer la dynamique du système complexe afin de mesurer dans chaque partition les éventuelles informations dépendant du temps. Pour cela, nous utiliserons une fenêtre glissante de longueur donnée L se déplaçant d'un état vers un autre état selon un pas unitaire commençant par la première valeur de x_0 et terminant vers la valeur x_{N-1} .

On définit alors $S_j^{(L)} = \{S_0^{(L)}, S_1^{(L)}, \dots, S_R^{(L)}\}$ avec $S_j^{(L)}$ correspondant au segment j de taille respective L et $j = 0, 1, \dots, R$ avec R correspondant au nombre total de segments S_j pour une certaine longueur L . Par exemple, si on choisit $L = 3$, alors on obtient :

$$\begin{aligned}
 S_j^{(L=3)} &= \{S_0^{(3)}, S_1^{(3)}, \dots, S_R^{(3)}\} \text{ avec :} \\
 S_0^{(3)} &= \{x_0, x_1, x_2\} \\
 S_1^{(3)} &= \{x_2, x_3, x_4\}, \\
 &\dots \\
 S_R^{(3)} &= \{x_{N-3}, x_{N-2}, x_{N-1}\}.
 \end{aligned}$$

Pour une question de lisibilité on définit :

$$S_j^{(L)} = \{x_{j,l}\}^{(L)} \quad (2.51),$$

Avec $j = 0, \dots, R$ la j -ème partition et $l = 1, 2, \dots, L$ avec L la taille de chaque partition. Ainsi, pour $L = 3$, on a : $S_j^{(3)} = \{x_{0,1}, x_{0,2}, x_{0,3}\}, \dots, \{x_{R,1}, x_{R,2}, x_{R,3}\}$. Parce que chaque symbole $x_{j,l}$ ne peut prendre que deux valeurs différentes 0 ou 1, par conséquent le nombre de partitions possibles S pour une taille L donnée est défini par la relation suivante :

$$\Phi(L) = 2^L \quad (2.52).$$

Ainsi, si on considère le même exemple que précédemment à savoir pour $L = 3$ alors $\Phi(L = 3) = 8$ et on obtient :

$$S^{(L=3)} = [\{0\ 0\ 0\}, \{0\ 0\ 1\}, \{0\ 1\ 1\}, \{1\ 1\ 1\}, \{1\ 1\ 0\}, \{1\ 0\ 0\}, \{1\ 0\ 1\}, \{0\ 1\ 0\}].$$

Maintenant que nous avons divisé la séquence de symboles en un ensemble de partitions de longueur égale constituées de 0 et 1, nous avons différents états $\Phi(L)$ possibles, états que nous appelons également structures ou patterns. En revanche, nous ne nous intéressons pas à la forme ni à la nature de ces structures en elles-mêmes, mais plutôt à leur distribution de fréquence, à savoir à quelle fréquence sommes-nous susceptibles de rencontrer ces structures. De cette façon, le point principal est de résumer la série chronologique économique originale en un ensemble de fréquences relatives, puis d'utiliser cet ensemble pour mener une étude basée sur la théorie de l'information. On peut facilement acquérir les probabilités de chaque partition en calculant les fréquences de différentes séquences de symboles qui se présentent comme suit :

$$p_j^{(L)} = \frac{\phi(S^{(L)})}{R} \quad (2.53),$$

avec $j = 1, \dots, R$, et $\phi(S^{(L)})$ est le nombre de fois qu'un segment S de taille L apparaît dans la série chronologique économique. De la même manière, si nous voulons connaître la probabilité de tous les segments possibles indépendamment de celui provenant des échelles de temps économiques elles-mêmes, dans ce cas nous définirons :

$$p_k^{(L)} = \frac{\phi^L}{R} \quad (2.54),$$

avec $k = 1, \dots, \Phi(L)$ où k est le nombre de séquences possibles de taille L et ϕ est le nombre de fois qu'une certaine séquence x_0, x_1, \dots, x_{N-1} apparaît.

Enfin, maintenant que nous avons une série de symboles divisée en un ensemble de partitions et que nous pouvons quantifier le degré d'information d'une partition basée sur la théorie de l'information -au sens que chaque partition est vue comme un événement avec une probabilité donnée-, nous souhaitons analyser ces différentes probabilités. Devant la difficulté de calculer ces probabilités quand on a affaire à des suites de séries de nombres binaires, nous proposons de convertir ces séries de nombres binaires en nombres fractionnaires. Définissons π un nombre fractionnaire associé à chaque partition donnée $S^{(L)}$ tel que :

$$\pi(S_j^{(L)}) = \sum_{l=1}^L x_{j,l} 2^{-l} \quad (2.55),$$

pour $j = 0, 1, \dots, R$ et $0 \leq \pi(S_j^{(L)}) < 1$. Encore une fois, si l'on prend en compte l'exemple ci-dessus avec $L = 3$, on obtient la nouvelle séquence de symboles suivante :

$$\pi(S^{(L=3)}) = \{0, 0.125, 0.375, 0.875, 0.75, 0.5, 0.625, 0.25\}.$$

Nous remplaçons chaque séquence de symboles (partition) par un nombre fractionnaire et unique compris entre 0 et 1, auquel nous avons attribué une probabilité donnée qui correspond à sa fréquence d'occurrence dans la série chronologique économique. La probabilité définie par l'équation (2.53) devient:

$$p_j^{(L)} = \frac{\phi(\pi(S^{(L)}))}{R} \quad (2.56),$$

avec $j = 1, \dots, R$, et $\phi(\pi(S^{(L)}))$ est le nombre de fois où un nombre fractionnaire $\pi(S^{(L)})$ d'une partition de taille L apparaît dans la série chronologique. Ce dernier fera l'objet d'une étude plus poussée dans la suite de la section.

2.5.2 *Preuve de présence de mémoire et de prédictibilité dans la série chronologique étudiée.*

Une manière qui permet de déduire la présence d'un processus de mémoire longue dans une séquence de symboles consiste à étudier le caractère markovien ou non-markovien de la série. La propriété de Markov indique que la distribution de probabilité conditionnelle pour le système étudié à l'étape suivante dépend uniquement de l'état actuel du système et est indépendant de ses états précédents. En considérant la série de suites de nombres fractionnaires définie ci-dessus par l'équation (2.55), si la suite $\pi(S_j^{(L)})$ est définie par un processus de Markov, on devrait alors avoir :

$$\begin{aligned} & P\left(\pi(S_{j+1}^{(L)}) \mid \pi(S_j^{(L)}), \pi(S_{j-1}^{(L)}), \dots, \pi(S_0^{(L)})\right) \\ &= P\left(\pi(S_{j+1}^{(L)}) \mid \pi(S_j^{(L)})\right) \quad (2.57). \end{aligned}$$

Avant d'exposer les probabilités conditionnelles de chaque nombre fractionnaire $\pi(S_j^{(L)})$, compte tenu des collections de sous-ensembles, nous proposons de comparer les probabilités des séquences de symboles empiriques avec des séquences parfaitement aléatoires pour une longueur de partition donnée $L = \{2, 4, 6, 8\}$, et de tester leur stationnarité. On montrera par la même occasion que l'analyse des probabilités conditionnelles est une méthode beaucoup plus fiable que les tests stationnaires lorsque nous voulons détecter le processus de mémoire dans une série de séquences de symbole.

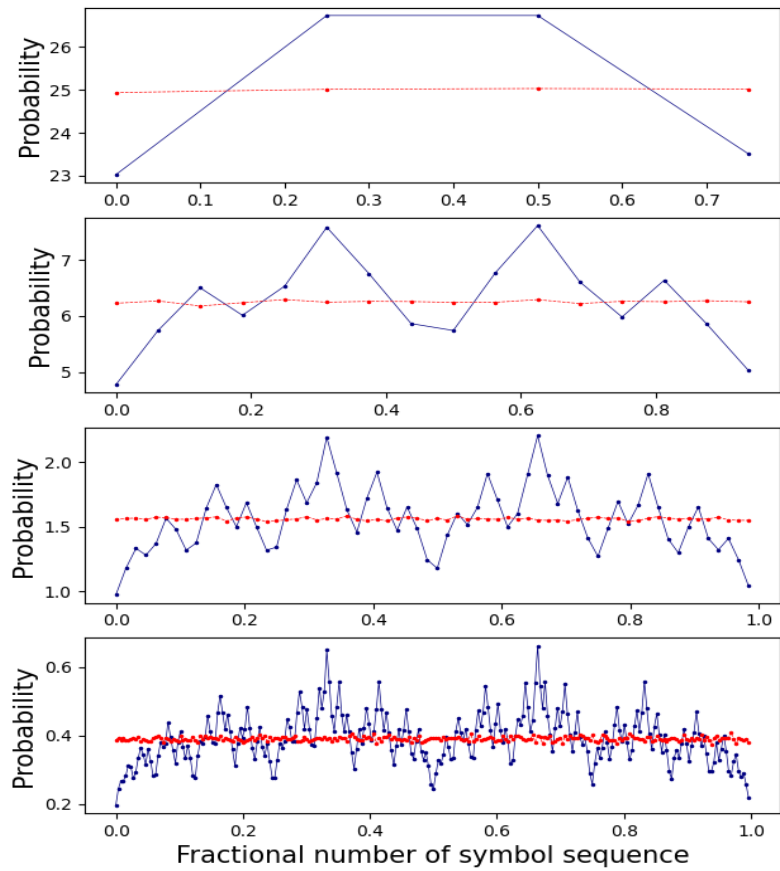


Figure 2. 6 Probabilités des rendements logarithmiques de la série chronologique sur l'échelle minute et pour les partitions de longueur $L = [2,4,6,8]$. Par conséquent, chaque série a respectivement 4, 16, 64 et 256 types de symboles possibles. En bleu, nous représentons les séquences de symboles dynamiques du Standard & Poor500, en rouge nous représentons une séquence de symboles dynamiques aléatoires.

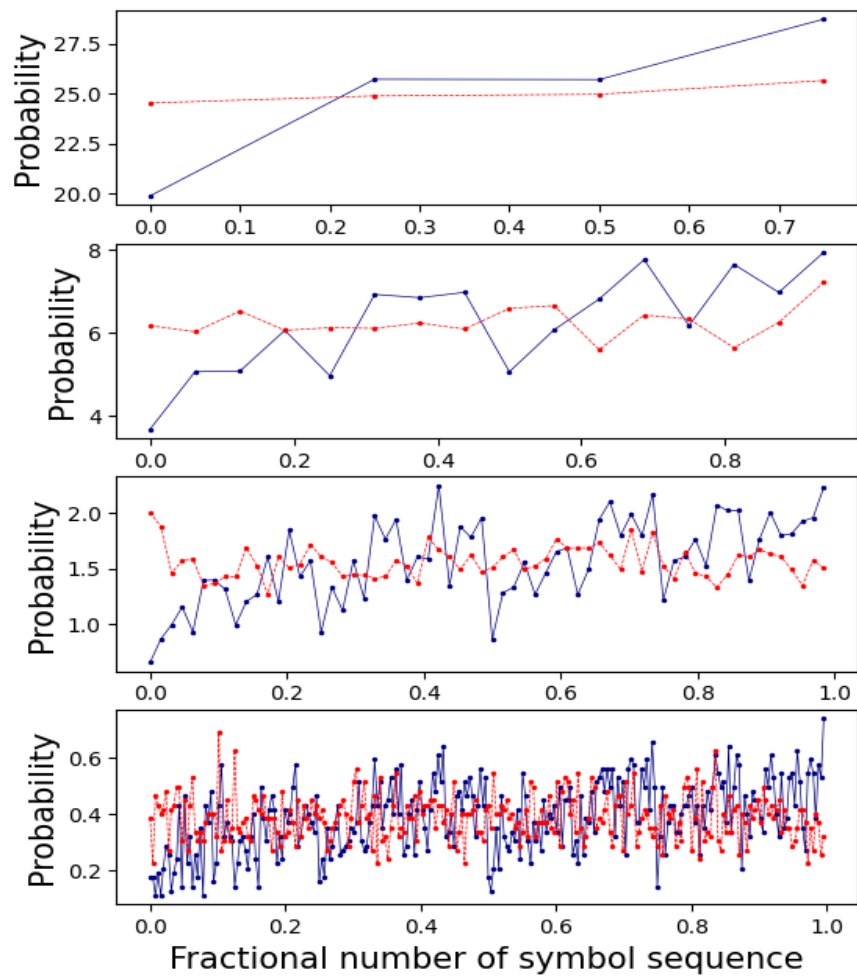


Figure 2. 7 Probabilités des rendements logarithmiques de la série chronologique sur l'échelle 4-hours et pour les partitions de longueur $L = [2,4,6,8]$. Par conséquent, chaque série a respectivement 4, 16, 64 et 256 types de symboles possibles. En bleu, nous représentons les séquences de symboles dynamiques du Standard & Poor500, en rouge nous représentons une séquence de symboles dynamiques aléatoires.

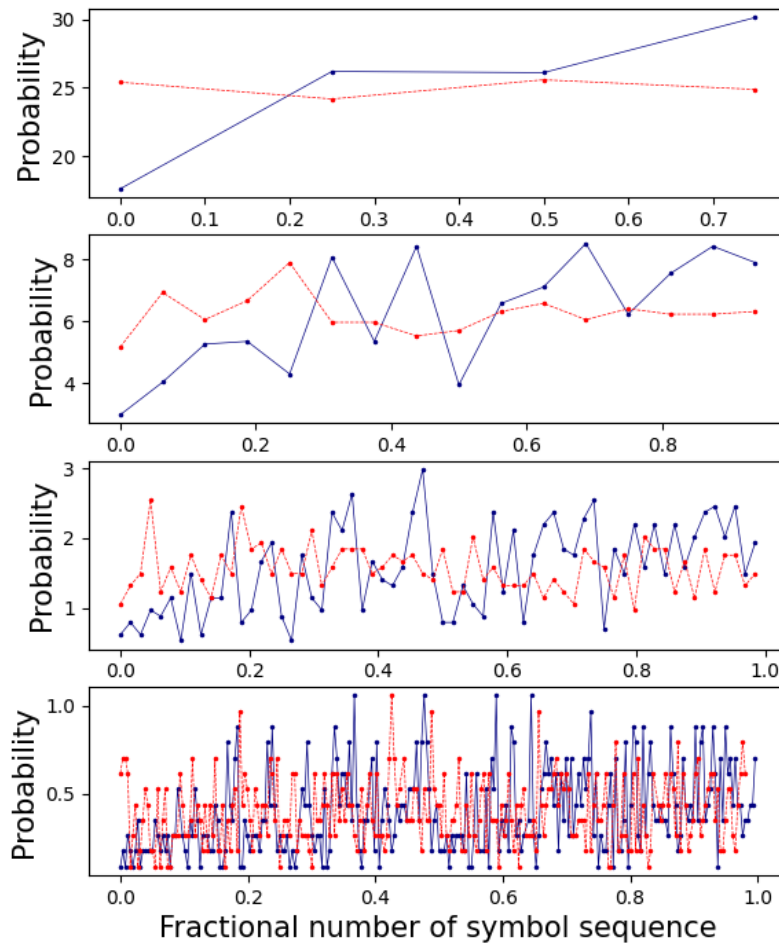


Figure 2. 8 Probabilités des rendements logarithmiques de la série chronologique sur l'échelle journalière et pour les partitions de longueur $L = [2,4,6,8]$. Par conséquent, chaque série a respectivement 4, 16, 64 et 256 types de symboles possibles. En bleu, nous représentons les séquences de symboles dynamiques du Standard & Poor500, en rouge nous représentons une séquence de symboles dynamiques aléatoires.

Les figures 2.6, 2.7, 2.8 mettent en avant les probabilités de la dynamique des rendements logarithmiques du Standard & Poor500 pour des séquences de symboles de différentes longueurs L par rapport à une séquence de symboles dynamiques aléatoires de la même longueur, sur différentes échelles de temps respectives, à savoir minute, 4-heures et journalière. Les différences de comportements entre la séquence de symboles de rendements logarithmiques et la séquence de symboles aléatoires sont observables sur chaque échelle de temps. Sur l'échelle de temps minute, nous constatons l'émergence de structures hiérarchiques organisées qui se répètent de manière cyclique, alors que le comportement de la série aléatoire semble rester parfaitement stationnaire et ne présente aucune caractéristique structurelle. Sur des échelles de temps plus grandes (c'est-à-dire 4-heures et journalière), les probabilités des séquences de symboles du Standard & Poor500 sont toujours différentes des séquences de symboles aléatoires, mais cette différence est moins évidente et nous observons moins d'émergence de structures et de motifs ordonnés. Lorsque la longueur L augmente, les séquences de symboles provenant d'échelles de temps plus grandes présentent des comportements moins distinguables d'un processus purement aléatoire. Cependant, on observe toujours une tendance de fonds haussière dans la dynamique des séquences de symboles laissant penser à une certaine dépendance de probabilités conditionnelles avec des états (rendement logarithmique positif ou négatif) précédents.

Dans le cadre de la théorie de l'information, nous abordons la question du processus de mémoire par la question qualitative suivante : la connaissance du passé réduit-elle notre incertitude sur l'avenir ?

Ainsi, nous préférons aborder la question de la prédictibilité au travers d'une analyse des probabilités conditionnelles des séquences de symboles défini comme suit : $P\left(\pi\left(S^{(L)}_{j+1}\right)\left|\pi\left(S^{(L)}_j\right),\pi\left(S^{(L)}_{j-1}\right),\dots,\pi\left(S^{(L)}_0\right)\right.\right)$. On nomme Ω l'espace d'échantillon tel que $\Omega = \left\{\left(S^{(L)}_j\right),\pi\left(S^{(L)}_{j-1}\right),\dots,\pi\left(S^{(L)}_0\right)\right\}$, \mathcal{F} la filtration représentant un sous-ensemble d'événements de Ω et P la probabilité de chaque

événement et les probabilités conditionnelles des séquences de symboles doivent satisfaire la contrainte suivante :

$$\sum_{S^{(L)} \in \Omega} P(S^{(L)} | \mathcal{F}) = 1 \quad (2.58).$$

La probabilité conditionnelle est un indicateur crucial dans la discussion de la propriété markovienne et, d'après la définition des chaînes de Markov en temps discret, on a :

$$\begin{aligned} \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\ = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \quad (2.59), \end{aligned}$$

$$\text{avec } \mathbb{P}(X_n | X_{n-1}) = \frac{\mathbb{P}(X_n, \dots, X_0)}{\mathbb{P}(X_{n-1}, \dots, X_0)} \quad (2.60).$$

Si la propriété markovienne est satisfaite et si on choisit de prendre le même exemple que précédemment, à savoir une séquence de symboles de longueur $L = 3$, alors nous devons trouver que :

$$\begin{aligned} p(x_{0,3} = 0 | x_{0,2} = 0, x_{0,1} = 0) \\ = p(x_{0,3} = 0 | x_{0,2} = 0, x_{0,1} = 1) \\ = p(x_{0,3} = 0 | x_{0,2} = 1, x_{0,1} = 1) \\ = p(x_{0,3} = 0 | x_{0,2} = 1, x_{0,1} = 0) \\ \text{et:} \\ p(x_{0,3} = 1 | x_{0,2} = 0, x_{0,1} = 0) \\ = p(x_{0,3} = 1 | x_{0,2} = 0, x_{0,1} = 1) \\ = p(x_{0,3} = 1 | x_{0,2} = 1, x_{0,1} = 1) \\ = p(x_{0,3} = 1 | x_{0,2} = 1, x_{0,1} = 0). \end{aligned}$$

En d'autres termes, si l'occurrence de séquences de symboles de taille L ne dépend que de l'état présent, et non de la séquence d'événements qui l'ont précédée, alors nous pouvons en déduire que le système résulte d'un processus markovien. Nous décidons d'étudier ces propriétés et nous présentons les résultats dans les tableaux ci-dessous.

	$p(0 0,★)$	$p(0 1,★)$	$p(1 0,★)$	$p(1 1,★)$
.	23.0264%	26.7324%	26.7324%	23.5089%
0,0	4.7725%	6.5309%	5.7409%	5.9821%
0,1	5.7409%	7.5822%	6.7721%	6.6371%
1,0	6.5011%	6.7556%	7.6120%	5.8636%
1,1	6.0119%	5.8636%	6.6073%	5.0261%
0,0,0,0	0.9751%	1.3436%	1.1806%	1.2732%
0,0,0,1	1.1806%	1.6356%	1.4363%	1.4884%
0,0,1,0	1.3341%	1.8640%	1.6054%	1.6977%
0,0,1,1	1.2828%	1.6877%	1.5186%	1.5228%
0,1,0,0	1.3720%	1.8413%	1.6498%	1.6677%
0,1,0,1	1.5675%	2.1923%	1.9118%	1.9106%
0,1,1,0	1.4801%	1.9149%	1.7088%	1.6518%
0,1,1,1	1.3213%	1.6337%	1.5017%	1.4070%
1,0,0,0	1.3760%	1.4608%	1.6033%	1.3008%
1,0,0,1	1.6459%	1.7229%	1.9057%	1.4976%
1,0,1,0	1.8280%	1.9239%	2.2057%	1.6544%
1,0,1,1	1.6513%	1.6480%	1.8972%	1.4107%
1,1,0,0	1.5018%	1.4746%	1.6818%	1.3239%
1,1,0,1	1.6871%	1.6538%	1.8848%	1.4114%
1,1,1,0	1.5016%	1.4882%	1.6268%	1.2471%
1,1,1,1	1.3214%	1.2471%	1.4139%	1.0438%

Tableau 2. 1 Probabilités conditionnelles des séquences de symboles de longueur $L = \{2,4,6\}$ sur l'échelle de temps minute.

	$p(0 0,★)$	$p(0 1,★)$	$p(1 0,★)$	$p(1 1,★)$
.	19.8812%	25.7106%	25.6946%	28.7137%
0,0	3.6787%	4.9639%	5.0602%	6.1687%
0,1	5.0602%	6.9237%	6.0723%	7.6466%
1,0	5.0763%	6.8434%	6.8112%	6.9719%
1,1	6.0562%	6.9719%	7.7590%	7.9357%
0,0,0,0	0.6588%	0.9320%	0.8677%	1.2213%
0,0,0,1	0.8677%	1.3338%	1.2856%	1.5748%
0,0,1,0	0.9963%	1.1249%	1.3338%	1.6069%
0,0,1,1	1.1570%	1.5748%	1.5587%	1.7676%
0,1,0,0	0.9320%	1.2373%	1.2695%	1.5266%
0,1,0,1	1.3980%	1.9765%	1.4623%	2.0730%
0,1,1,0	1.3980%	1.7676%	1.6552%	2.0247%
0,1,1,1	1.3177%	1.9444%	1.6873%	2.0247%
1,0,0,0	0.9963%	1.3980%	1.2695%	1.3980%
1,0,0,1	1.2052%	1.6069%	1.4945%	1.7676%
1,0,1,0	1.2695%	1.5909%	1.9444%	2.0087%
1,0,1,1	1.6069%	2.2497%	2.1051%	1.7998%
1,1,0,0	1.2052%	1.3498%	1.7998%	1.8158%
1,1,0,1	1.8480%	1.8801%	1.9926%	1.9283%
1,1,1,0	1.4302%	1.7837%	1.7998%	1.9605%
1,1,1,1	1.5748%	1.9605%	2.1694%	2.2336%

Tableau 2. 2 Probabilités conditionnelles des séquences de symboles de longueur $L = \{2,4,6\}$ sur l'échelle de temps 4-hours.

	$p(0 0,★)$	$p(0 1,★)$	$p(1 0,★)$	$p(1 1,★)$
.	17.6007%	26.1821%	26.0946%	30.1226%
0,0	2.9825%	4.2982%	3.9474%	6.2281%
0,1	4.0351%	8.0702%	6.5789%	7.5439%
1,0	5.2632%	5.3509%	7.1053%	8.4211%
1,1	5.3509%	8.4211%	8.5088%	7.8947%
0,0,0,0	0.6151%	0.8787%	0.7909%	0.7030%
0,0,0,1	0.7909%	0.5272%	0.7909%	1.8453%
0,0,1,0	0.6151%	1.7575%	1.3181%	1.4938%
0,0,1,1	0.9666%	1.1424%	1.0545%	2.1968%
0,1,0,0	0.8787%	0.9666%	0.8787%	1.5817%
0,1,0,1	1.1424%	2.3726%	2.3726%	2.1968%
0,1,1,0	0.5272%	2.1090%	1.2302%	1.4938%
0,1,1,1	1.4938%	2.6362%	2.1090%	2.1968%
1,0,0,0	0.6151%	0.9666%	0.7909%	1.5817%
1,0,0,1	1.1424%	1.6696%	1.7575%	2.0211%
1,0,1,0	1.1424%	1.4060%	2.1968%	2.3726%
1,0,1,1	2.3726%	1.3181%	2.3726%	2.4605%
1,1,0,0	0.7909%	1.5817%	1.8453%	2.0211%
1,1,0,1	0.9666%	2.3726%	1.7575%	2.4605%
1,1,1,0	1.6696%	2.9877%	2.2847%	1.4938%
1,1,1,1	1.9332%	1.4938%	2.5483%	1.9332%

Tableau 2. 3 Probabilités conditionnelles des séquences de symboles de longueur $L = \{2,4,6\}$ sur l'échelle de temps journalier.

Les tableaux 2.1, 2.2 et 2.3, présentent la liste des probabilités conditionnelles pour les échelles de temps respectives à savoir minutes, 4-heures et journalière selon une longueur de partition différente tel que $L = \{2,4,6\}$. Nous proposons d'étudier le comportement non-markovien [122] de la dynamique des rendements logarithmiques du Standard&Poor500 selon différentes échelles de temps et spatiales, (différente taille de séquences de symboles $L = \{2,4,6\}$) afin d'en déduire la présence ou non d'un processus de mémoire. Ainsi, si nous prenons comme exemple le tableau 2.1, la valeur de la colonne $p(0|0,★)$ représente la probabilité que l'événement suivant soit 0 (c'est-à-dire un rendement négatif) sachant que l'événement précédent était également 0 suivi d'une série de symboles représentés par $★$ qui correspondent à des séries de bits $\{0;1\}$ dont la taille dépend de la valeur correspondante L . Ainsi si $★$ correspond à la cellule du tableau qui contient la séquence de symboles $\{0,0,0,0\}$, cela signifie que la taille correspondante à la séquence est $L = 6$ et la probabilité conditionnelle que l'on analyse devient $p(0|0,0,0,0,0,0)$.

Nous analysons les résultats des tableaux 2.1, 2.2 & 2.3 (p.79, 80). Nous constatons que sur des échelles de temps petites (c.-à-d. minutes) le caractère non-markovien pour de petite longueur de séquence de symboles ($L = 2$) est observable puisque l'égalité parfaite entre les probabilités conditionnelles n'est pas parfaitement respectée avec par exemple $p(0|0) \neq p(1|1)$; cependant la réciproque est également vraie puisque nous observons que $p(0|1) = p(1|0)$. Cette observation signifie que nous ne pouvons pas facilement conclure que cette séquence de symboles est régie par un processus purement non-markovien. Par ailleurs, on observe une certaine dispersion (environ 1.7%) entre des probabilités conditionnelles sur cette même échelle de temps et toujours pour une taille $L = 2$, allant de 23% à 26,7%.

Par conséquent, sur de petites échelles de temps, la dynamique des rendements logarithmiques du Standard&Poor500 se caractérise par une dépendance entre l'événement futur et la valeur de l'évènement actuel traduisant une certaine prévisibilité du fait de son faible caractère non-Markovien.

Sur des échelles de temps plus longues (c.-à-d. 4heures et journalière), l'observation montre que, pour une même longueur de séquence de symboles $L = 2$, il y a une parfaite non-uniformité de probabilités conditionnelles, de sorte que $p(0|0) \neq p(1|1) \neq p(1|0) \neq p(0|1)$. Cela implique que, sur des échelles de temps plus longues, la dynamique des rendements logarithmiques Standard&Poor500 présente également un fort caractère non markovien avec une dépendance aux événements précédents. De plus, si $p(1|0)$ est légèrement différent de $p(0|1)$ la dispersion des probabilités est suffisamment significative (3.2% pour la séquence de symboles sur l'échelle 4-heures et 4.57% pour la séquence de symboles sur l'échelle journalière) pour en déduire une inégalité parfaite de probabilités conditionnelles.

On observe également que la variabilité des probabilités conditionnelles augmente à mesure que les échelles de temps deviennent plus grandes, et nous constatons également que la probabilité conditionnelle maximale pour les échelles de

temps 4-heures et journalière est atteinte pour $p(1|1)$. En d'autres termes, la probabilité que l'événement suivant soit un rendement positif est maximale lorsque l'événement précédent est un rendement positif, ce qui est cohérent avec la croissance haussière à long terme du Standard & Poor500 sur la période étudiée et crédibilise l'observation de tendance de fonds haussière dans la série des probabilités conditionnelles observées lors de l'analyse des figures 2.7 et 2.8 (p.75, 76).

Pour les autres longueurs de séquences de symboles, à savoir $L = \{4, 6, 8\}$, nous analysons les résultats à travers des graphiques à moustaches afin de traiter des statistiques descriptives, ce qui est plus pratique que les tableaux 2.1, 2.2 & 2.3 (p.79, 80) en raison de la plus grande taille des séquences de symboles. On rappelle que pour $L = 8$, on a $\Phi(L = 8) = 2^8 = 256$ séquences de symboles possibles.

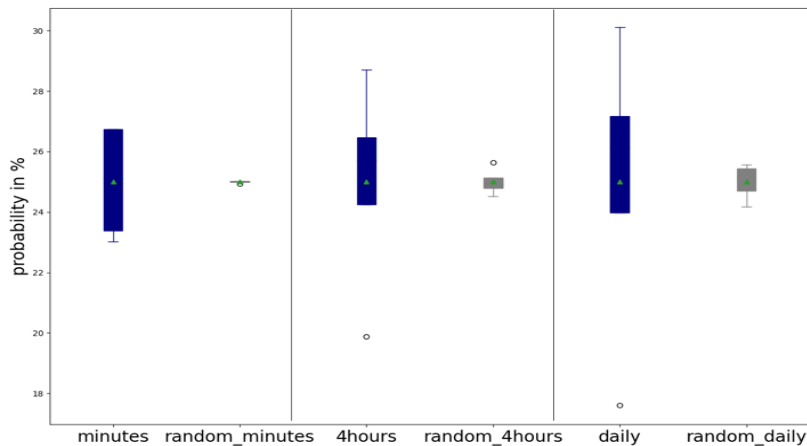


Figure 2.9 En bleu, graphique Box-Plot représentant les caractéristiques statistiques des séquences de symboles de longueur $L = 2$ sur différentes échelles de temps (minutes, 4 heures, par jour) de la dynamique des retours logarithmiques de Standard & Poor500. En gris, graphique Box-Plot représentant les caractéristiques statistiques de séquences de symboles de longueur $L = 2$ sur différentes échelles de temps (minutes, 4 heures, par jour) de symboles dynamiques parfaitement aléatoires.

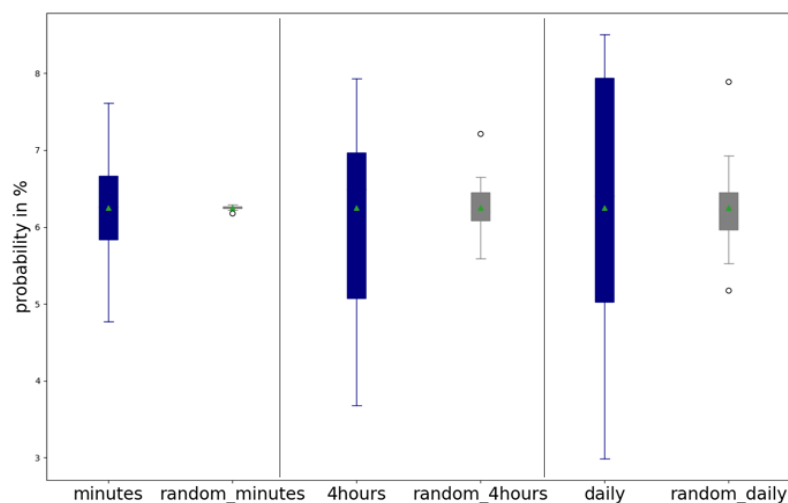


Figure 2. 10 En bleu, graphique Box-Plot représentant les caractéristiques statistiques des séquences de symboles de longueur $L = 4$ sur différentes échelles de temps (minutes, 4 heures, par jour) de la dynamique des retours logarithmiques de Standard & Poor500. En gris, graphique Box-Plot représentant les caractéristiques statistiques de séquences de symboles de longueur $L = 4$ sur différentes échelles de temps (minutes, 4 heures, par jour) de symboles dynamiques parfaitement aléatoires.

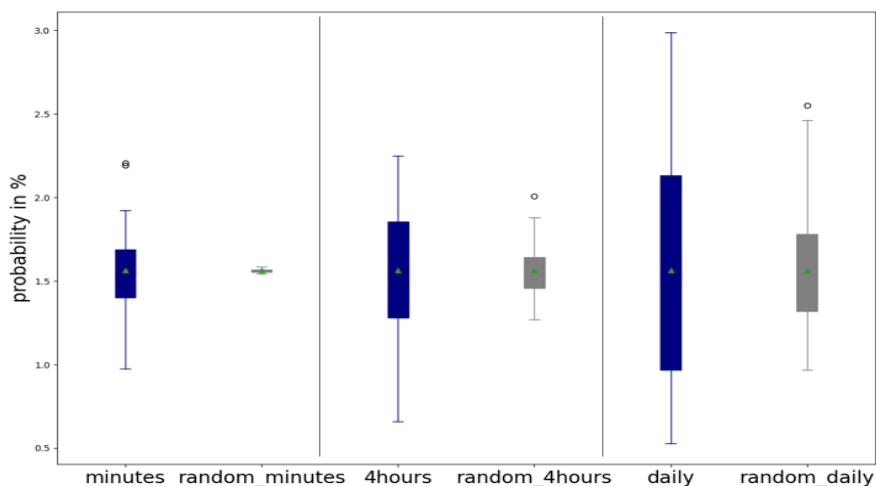


Figure 2. 11 En bleu, graphique Box-Plot représentant les caractéristiques statistiques des séquence de symboles de longueur $L = 6$ sur différentes échelles de temps (minutes, 4 heures, par jour) de la dynamique des retours logarithmiques de Standard & Poor500. En gris, graphique Box-Plot représentant les caractéristiques statistiques de séquences de symboles de longueur $L = 6$ sur différentes échelles de temps (minutes, 4 heures, par jour) de symboles dynamiques parfaitement aléatoires.

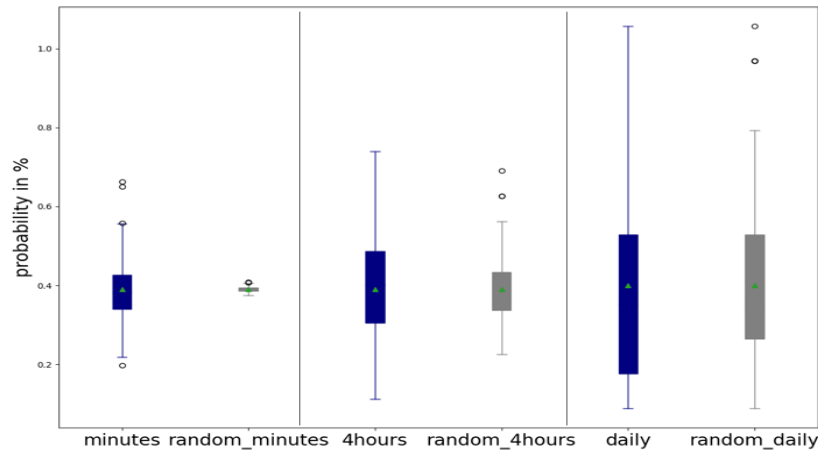


Figure 2. 12 En bleu, graphique Box-Plot représentant les caractéristiques statistiques des séquences de symboles de longueur $L = 8$ sur différentes échelles de temps (minutes, 4 heures, par jour) de la dynamique des retours logarithmiques de Standard & Poor500. En gris, graphique Box-Plot représentant les caractéristiques statistiques de séquences de symboles de longueur $L = 8$ sur différentes échelles de temps (minutes, 4 heures, par jour) de symboles dynamiques parfaitement aléatoires.

Les figures 2.9, 2.10, 2.11 et 2.12 mettent en évidence une série de graphiques de moustaches pour différentes séquences de symboles respectives, longueur $L = \{2,4,6,8\}$. En bleu, nous représentons les caractéristiques statistiques de la séquence de symboles à partir de la dynamique des rendements logarithmiques Standard & Poor500, et nous les comparons avec les statistiques provenant de séquences de symboles de même longueur mais purement aléatoires.

Lorsque $L = 2$, sur des échelles de temps plus petites (minutes), les probabilités conditionnelles sont beaucoup plus dispersées sur la dynamique des symboles Standard & Poor500 que sur la dynamique des symboles aléatoires avec une taille d'interquartile beaucoup plus large. La taille des interquartiles permettant de mieux comparer la dispersion des probabilités issues de la dynamique du S&P500 et ceux issus de séries aléatoires est exposée dans le tableau 2.4. On y montre notamment le rapport entre la longueur de l'interquartile de la dynamique Standard & Poor500 et la longueur de l'interquartile de la dynamique aléatoire. Nous constatons par exemple que, sur des échelles de temps petites (minute) et pour une faible taille de séquence de

symboles ($L = 2$), la longueur de l'interquartile du S&P500 est 122 fois plus grande que celle d'une série aléatoire, preuve d'une forte dispersion et non uniformité de probabilités conditionnelles.

Ratio \emptyset	L = 2	L = 4	L = 6	L = 8
Minutes	122.7	36	20.12	10.96
4hours	6.51	5.14	3.11	1.88
Daily	4.29	6.05	2.52	1.33

Tableau 2. 4 Comparaison pour différentes longueurs de séquence L et pour différentes échelles de temps. Le rapport de longueur interquartile (IQR) que nous définissons comme suit : $\emptyset = \frac{\text{length of IQR from SP500}}{\text{length IQR from random series}}$.

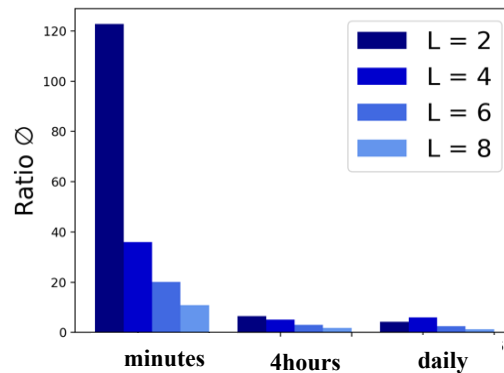


Figure 2. 13 Histogramme du ratio \emptyset selon les échelles de temps (minute, 4-hours, journalière) et selon l'échelle spatiale des séquences de symboles ($L = \{2,4,6,8\}$).

Lorsque $L = \{4,6,8\}$, nous observons toujours pour les échelles de temps de 4 heures et journalière une légère différence entre leur médiane respective et celle de la séquence de symboles aléatoires. De plus, sur toutes les échelles de temps, il est évident que la forte dispersion des probabilités conditionnelles de la dynamique Standard&Poor500 persiste et que la longueur de leur interquartile respectif est beaucoup plus longue que celle de la séquence de symboles aléatoires.

Cependant, cette différence de longueur interquartile diminue à mesure que la longueur L de la séquence de symboles augmente. Ceci est observable à la fois dans le tableau 2.4 (p.85) et la figure 2.13 (p.85) ; pour chaque échelle de temps, au fur et à mesure que nous augmentons le nombre d'événements passés, le Standard & Poor500 a tendance à converger plus rapidement vers des statistiques similaires de celles purement aléatoires, bien que pour $L = 8$, la dynamique des séquences de symboles reste fortement non-markovienne. Néanmoins, la dépendance d'un événement futur à une série d'événements passés diminue à mesure que la longueur des données historiques augmente. Et cette décroissance est plus rapide sur des échelles de temps plus petites que sur des échelles plus longues (tableau 2.4, p.85).

En résumé, au travers d'une analyse de séquences de symboles traduisant la dynamique des rendements logarithmiques du S&P500, on comprend que la prévisibilité des rendements va dépendre de :

- L'échelle de temps : sur des échelles de temps petites (minutes), la dispersion des probabilités conditionnelles des séquences de symboles se distingue de manière plus significative des séquences de symboles purement aléatoires que sur des échelles de temps plus grandes. Nous en déduisons que sur, ces échelles petites, l'occurrence d'événements futurs nous transmet une information plus prévisible au sens de Shannon, c'est-à-dire que notre effet de surprise sur l'observation future de rendements est amoindri. Par conséquent, sur ces échelles petites, nous aurons tendance à constater moins de désordre et moins d'incertitude.
- La taille des événements historiques considérés : pour chaque échelle de temps, la dépendance des événements futurs décroît à mesure que la taille des événements historiques augmente. Et cette dépendance à la taille est plus significative pour des échelles de temps petites (minutes).

2.5.3 Entropie et mesure de la complexité

Dans la section précédente, nous avons observé et décrit le caractère prévisible du S&P500 au travers d'une analyse non-markovienne de sa dynamique symbolique. Nous souhaitons maintenant mesurer le degré de prévisibilité des rendements logarithmiques futurs conditionnés aux rendements logarithmiques historiques en mesurant notre degré d'incertitude sur l'occurrence des rendements futurs toujours en fonction de la taille de l'historique considéré. Plus les probabilités conditionnelles d'événements futurs (0 ou 1), conditionnés à une série d'événements passés (série historique d'une certaine longueur) sont diverses, plus faible sera notre incertitude sur l'événement futur ; on parle ici d'incertitude au sens de Shannon.

Nous considérons le degré de prévisibilité comme étant une notion associée au degré d'incertitude. Nous rappelons l'information de Shannon comme étant étroitement associée à l'incertitude selon l'équation (2.13). Nous en donnons une relation modifiée pour tenir en compte de l'échelle spatiale L :

$$I(r^{(L)}) = \sum_i p_i^{(L)} \ln p_i^{(L)} \quad (2.61),$$

avec I l'Information de Shannon correspondant à la fonction de distribution de tout événement possible i , $r^{(L)} = \Phi(L)^{-1} = 2^{-L}$ la taille d'échelle de la série et p_i la probabilité de cet événement.

On s'intéresse tout particulièrement à la mesure négative de l'Information aussi appelée entropie de Shannon donnée par l'équation (2.21). Nous en donnons une relation modifiée pour tenir en compte de l'échelle spatiale L :

$$S(r^{(L)}) = -I(r^{(L)}) \quad (2.62).$$

Une séquence de symboles à faible entropie aura tendance à exhiber certaines formes géométriques telles que des motifs, des structures ordonnées et complexes que l'on considère comme étant plus ou moins prévisibles, tandis qu'une séquence de symboles à forte entropie aura tendance à être uniformément distribuée et parfaitement aléatoire.

Par conséquent, nous pouvons mesurer la prévisibilité et le processus de mémoire du Standard&Poor500 en déterminant l'entropie du système.

A cet effet, nous proposons d'appliquer l'Information Généralisée aussi appelée Information de Rényi [121] car elle permet de généraliser l'information de Shannon en tenant compte d'un nombre arbitraire et non uniforme appelé q . Ce dernier permet d'attribuer un degré d'importance à la probabilité d'un événement ; le but est de caractériser le degré de complexité d'une séquence de symboles en calculant un spectre complet d'informations généralisées (c.-à-d. l'entropie).

L'information généralisée de Rényi est donnée par équation (2.23). Nous en donnons une forme modifiée afin de tenir compte de l'échelle spatiale L comme suit :

$$I(r^{(L)})_q = \frac{1}{q-1} \log \sum_i^{\Phi(L)} (p_i^{(L)})^q \quad (2.63)$$

Ainsi, afin de mesurer le niveau de complexité du Standard&Poor500, nous analysons dans un premier temps l'évolution de l'information de Rényi selon différentes échelles spatiotemporelles, à savoir selon la valeur inverse de la longueur des séquences de symboles que nous avons définie par $r^{(L)}$; puis, dans un deuxième temps, nous mesurons le changement de l'information négative de Rényi selon différentes échelles de temps lorsque la longueur L des séquences de symboles augmente.

Nous avons besoin de la dimension Rényi définie par équation (2.23) qui mesure l'évolution de l'information Rényi définie par la relation suivante. Nous modifions l'équation (2.23) pour tenir compte des différentes échelles spatiales L , ce qui donne :

$$D_q = \lim_{r^{(L)} \rightarrow 0} \frac{I(r^{(L)})_q}{\ln r^{(L)}} \quad (2.64),$$

$$D_q = \lim_{r^{(L)} \rightarrow 0} \frac{1}{\ln r^{(L)}} \frac{1}{q-1} \log \sum_i^{\Phi(L)} (p_i^{(L)})^q \quad (2.65)$$

La dimension de Rényi est une fonction décroissante de q et restera finie tant que $r^{(L)} \rightarrow 0$.

Nous avons également besoin de l'entropie de Rényi définie par équation (2.49) qui correspond à l'information négative de Rényi. Nous modifions l'équation (2.49) pour tenir compte de l'échelle spatiale L , comme suit :

$$S_q = \lim_{L \rightarrow \infty} \frac{-I(r^{(L)})_q}{L} \quad (2.66)$$

$$S_q = \lim_{L \rightarrow \infty} \frac{1}{1-q} * \frac{1}{L} * \log \sum_i^{\Phi(L)} (p_i^{(L)})^q \quad (2.67)$$

L'entropie de Rényi à la même dépendance sur q que D_q . Nous proposons de mettre en évidence différentes entropies de Rényi pour différentes valeurs de q allant de -40 à 40, selon différentes échelles de temps pour différentes longueurs L et selon l'équation (2.67).

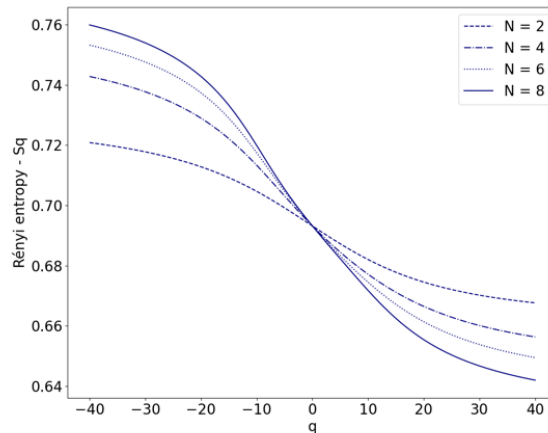


Figure 2. 14 Entropie de Rényi de la dynamique des séquences de symboles Standard&Poor500 pour $L = \{2,4,6,8\}$ selon l'échelle de temps minute.

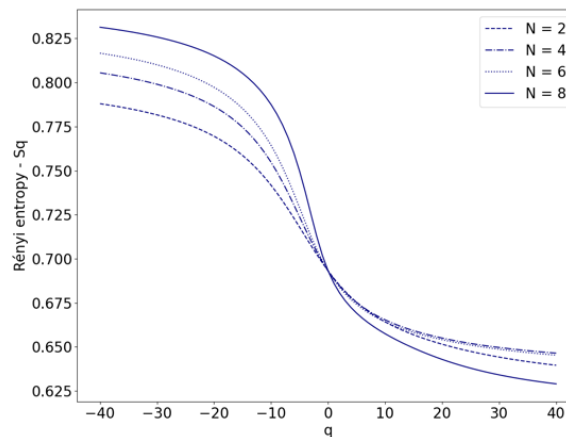


Figure 2. 15 Entropie de Rényi de la dynamique des séquences de symboles Standard&Poor500 pour $L = \{2,4,6,8\}$ selon l'échelle de temps 4-heures.

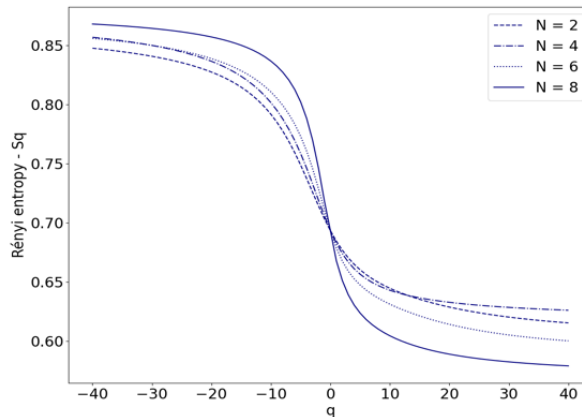


Figure 2. 16 Entropie de Rényi de la dynamique des séquences de symboles Standard&Poor500 pour $L = \{2,4,6,8\}$ selon l'échelle de temps journalière.

Les figures 2.14, 2.15 et 2.16 mettent en évidence les informations négatives de Rényi et leur dépendance vis-à-vis de q . Nous observons que, pour chaque échelle de temps, nous avons une forte dépendance des entropies sur q avec des décroissances relativement fortes et rapides à mesure que q augmente. En outre, plus la longueur de la séquence de symboles est grande, plus cette dépendance est forte, ce qui signifie que la complexité du Standard&Poor500 augmente à mesure que le nombre d'événements historiques pris en compte augmente.

Nous observons également que les entropies de Rényi ont une plus forte dépendance sur q selon les échelles de temps plus grandes que des échelles de temps plus petites. Cette observation est mise en évidence par la figure 2.14, à la fois pour $L = 2$ et $L = 8$, l'entropie de Rényi est beaucoup plus plate et son évolution plus monotone sur l'échelle de temps minute que sur l'échelle de temps journalière, preuve d'une dépendance des probabilités conditionnelles plus forte sur q pour de larges échelles temporelles.

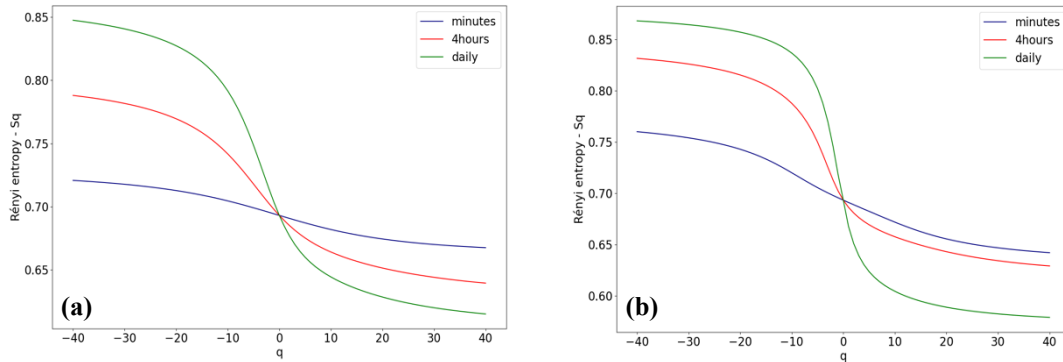


Figure 2.17 Comparaison des entropies de Rényi de la dynamique des séquences de symboles Standard & Poor500 pour $L = 2$ (a) et pour $L = 8$ (b) selon différentes échelles de temps.

Nous observons également à partir de la figure 2.17 que pour les valeurs négatives de q , les entropies de Rényi sur des échelles de temps petites (minute) sont plus faibles que sur des échelles de temps plus longues. Nous en déduisons que la séquence de symboles sur des échelles de temps petite présente moins d'incertitude pour les valeurs négatives de q que sur des échelles de temps plus grandes.

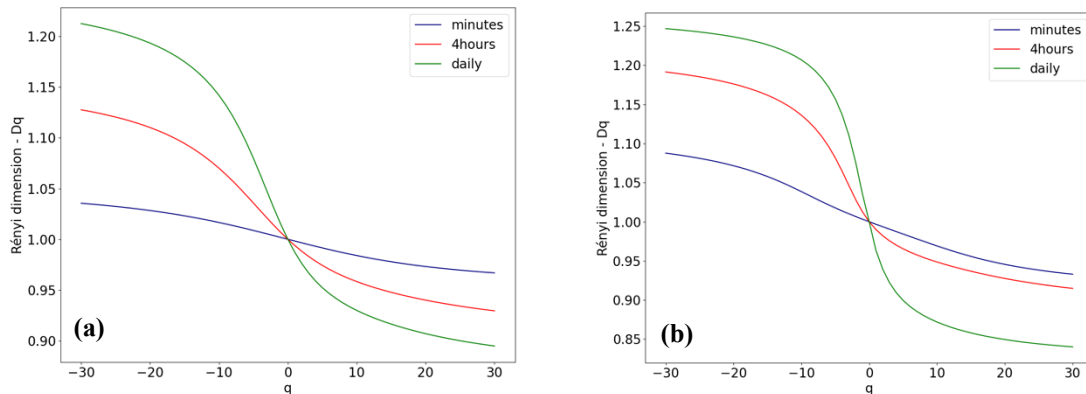


Figure 2.18 Comparaison des dimensions de Rényi de la dynamique des séquences de symboles Standard & Poor500 pour $L = 2$ (a) et pour $L = 8$ (b) selon différentes échelles de temps.

La figure 2.18 représente la dimension de Rényi selon la relation (2.65). On observe au travers de cette figure d'une part que plus la longueur L de la séquence de symboles est grande, plus la dépendance de la dimension à des valeurs différentes de

q est grande. D'autre part, plus les échelles de temps sont grandes, plus la dépendance sur q est également forte. On note également que sur l'échelle de temps minute et pour une petite longueur L , la dimension de Rényi est relativement plate et proche d'une dimension unitaire 1, mettant en évidence une propriété mono-fractale.

Par conséquent, nous pouvons en déduire que la complexité du Standard & Poor500 et de sa propriété de multi-fractalité augmente à mesure que la taille des échelles de temps et que la taille des séquences de symboles augmente.

Nous proposons maintenant d'interpréter la dimension généralisée en introduisant le spectre de singularité $f(\alpha)$ introduit par Halsey [61] en 1986 et que nous avons explicité dans la section 3.3.2. On rappelle que pour calculer $f(\alpha(q))$, on utilise la méthode développée par Chhabra qui permet d'éviter la transformée de Legendre de $\tau(q)$ et définit le spectre de singularité selon les équations (2.47) et (2.48).

Pour rappel, les écarts de valeurs de α indiquent les variétés d'échelle présentes dans la séquence de symboles alors que $f(\alpha)$ indique la force de contribution de chaque α .

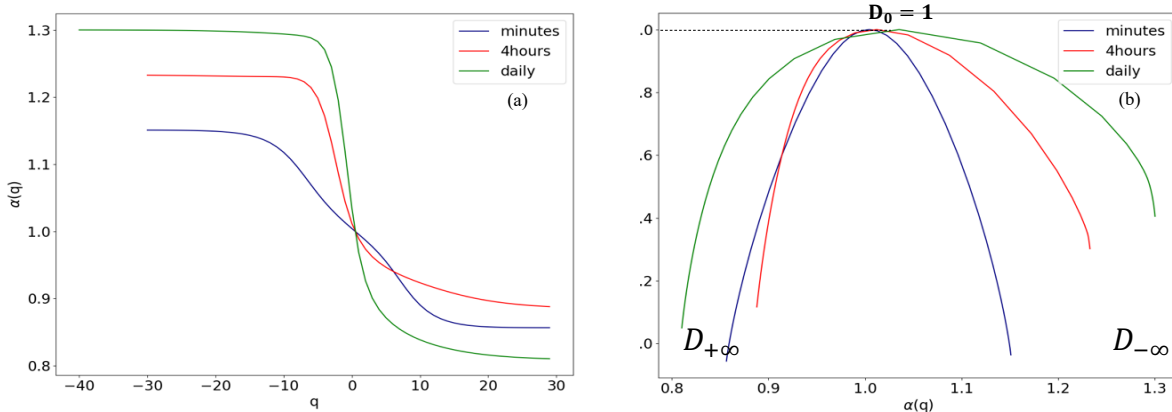


Figure 2. 19 (a) Comparaison des courbes $\alpha(q)$ versus q définies par l'équation (3.19) selon différentes échelles de temps ; (b) comparaison des fonctions $f(\alpha(q))$ contre $\alpha(q)$ pour les valeurs d'échelle spatiale allant de 2 à 8 et q allant de -30 à 30 selon différentes échelles de temps.

Figure 2.19 représente les différences structurelles de la dynamique du Standard&Poor500 selon différentes échelles de temps. La forte différence de valeur et de forme du spectre de singularité entre l'échelle de temps minute et celle journalière traduit le caractère multi-fractale du Standard&Poor500, à savoir que ce dernier décrit des statistiques différentes selon l'échelle spatio-temporelle.

L'exposant de singularité $\alpha(q)$ représente l'équilibre entre la dynamique des séquences de symboles de longueurs petites avec la dynamique de séquences de longueurs plus grandes. Nous observons que pour des valeurs q positives, les $\alpha(q)$ sont relativement plus faibles et l'écart des $\alpha(q)$ selon les échelles de temps différentes diminue, ce qui signifie que le nombre de séquences de symboles avec de fortes probabilités n'est pas courant, et cela sur toutes les échelles de temps. Pour les valeurs négatives de q , nous observons des valeurs plus grandes de $\alpha(q)$ avec une augmentation de l'écart entre les valeurs de $\alpha(q)$ selon les échelles de temps respectives, confirmant le fait que les séquences de symboles sont principalement dominées par des événements à faible probabilités conditionnelles, et cela également sur toutes les échelles de temps.

Les courbes de spectre multi-fractales (Figure 2.19. (b)) représentent la distribution des séquences de symboles pour différents facteurs d'échelle. La partie droite de chaque courbe (liée à des valeurs de q négatives) s'aplatit et s'étale de plus en plus à mesure que l'échelle de temps s'élargit, de sorte que $f\alpha(q)$ prend des valeurs plus grandes à mesure que les échelles de temps augmentent. La partie gauche de chaque courbe (liée à des valeurs de q positives) est moins aplatie et décroît rapidement, notamment pour les petites échelles de temps de 4 heures et minute.

L'évolution du spectre de singularité est un indicateur de la façon dont la dynamique des rendements logarithmiques du Standard&Poor500 change et évolue en fonction de l'échelle de temps et de l'échelle spatiale d'étude, preuve là encore de la présence de statistiques différentes et du caractère multi-fractal de ce dernier.

Enfin, nous observons une symétrie presque parfaite de $f(\alpha(q))$ sur l'échelle de temps minute, ce qui signifie que, pour $q \in [-\infty; +\infty]$, les séquences de symboles sont assez similaires avec généralement moins de variétés de probabilités conditionnelles entre les symboles séquences, ce qui confirme nos analyses dans les sous-sections supérieures. Cependant, concernant les échelles de temps de 4-heures et journalière, nous observons une forte concentration de valeurs de $f(\alpha(q))$ lorsque $q \rightarrow -\infty$, ce qui signifie qu'il existe de plus grandes variétés de séquences de symboles avec un nombre de probabilités conditionnelles faibles majoritaires, et confirme également notre précédente observation.

2.6 Conclusion

Dans ce chapitre nous avons dans un premier temps mis en avant les différentes méthodes de mesures de complexités au travers de la géométrie fractale. Nous avons montré que les mesures géométriques, notamment la dimension d'Hausdorff, sont efficaces et faciles à implémenter pour des systèmes complexes dont la statistique des structures sous-jacentes est invariante aux échelles spatiales. Ce type de mesure assimile le caractère complexe d'un système aux propriétés géométriques non euclidiennes de ce dernier.

En revanche, ces mesures ne sont pas adaptées lorsque ces systèmes présentent des variances de structures. Pour ce faire, nous introduisons des méthodes de mesures statistiques, à savoir la dimension généralisée de Rényi et le spectre de singularité qui permettent de mesurer le niveau d'information du système. L'avantage de ces méthodes est que, non seulement elles permettent d'étudier le caractère complexe d'un système selon des espaces spatio-temporels différents, mais elles sont également relativement simples en termes de calcul lorsqu'on utilise la théorie de la dynamique symbolique permettant d'encoder des séries de nombres quelconques en séries de bits et de symboles.

Ces méthodes de mesures ont fait l'objet d'applications sur une série temporelle de rendements de l'indice boursier Standard&Poor500. Les caractéristiques de prévisibilité et de complexité du Standard&Poor500 ont été examinées selon différentes échelles temporelles et spatiales en considérant ce dernier comme étant un système du type « coarse-grained ». Pour ce faire, nous avons utilisé les notions de mesures de complexité précédentes, à savoir la méthode de la dynamique symbolique et la théorie de l'information pour encoder la série temporelle en partition de symboles discrets, et la notion d'entropie et de dimension généralisée comme mesure de la complexité.

Premièrement, nous avons mis en évidence les comportements non-markoviens de la dynamique des rendements logarithmiques du Standard&Poor500 en mettant en évidence des probabilités conditionnelles inégales entre des séquences de symboles de

même longueur et, ce, selon des échelles de temps différentes. Nous prouvons ainsi la présence de processus de mémoire et de dépendance de rendements futurs aux rendements historiques.

Nous avons également constaté que, si, sur des échelles de temps plus longues, les rendements logarithmiques positifs étaient plus susceptibles d'être suivis par des rendements logarithmiques positifs, sur de petites échelles de temps (minutes), les rendements logarithmiques positifs (négatifs) étaient plus susceptibles d'être suivis par des rendements logarithmiques négatifs (positifs). Nous avons montré également que sur des échelles de temps plus petites (minute), la dynamique des rendements logarithmiques du Standard&Poor500 faisait émerger des structures plus ordonnées et hiérarchisées que sur des échelles de temps plus grandes, preuve, au sens de la théorie de Shannon, que le dynamisme de ce dernier exhibe moins d'informations spécifiques (moins de surprise) lorsque de nouveaux rendements occurrents.

Deuxièmement, à partir de ces probabilités conditionnelles des séquences de symboles correspondantes calculées sur différentes échelles de temps, une entropie de Rényi et un spectre de dimension non triviaux ont été trouvés. Nous avons constaté une dépendance plus forte de l'entropie aux valeurs de q lorsque les échelles de temps devenaient plus grandes (4-heure et journalière), avec une forte décroissance de la valeur de la dimension D_q à mesure que q augmente, et, à l'inverse, une dimension relativement plate et monotone sur de petites échelles de temps (minute). Nous en déduisons que le Standard&Poor500 a des propriétés multi fractales sur des échelles de temps plus larges et mono-fractales sur des échelles plus petites.

Enfin, nous avons calculé le spectre de singularité multi-fractal. Des preuves similaires ont été trouvées avec plus d'homogénéité et de symétrie de $f(\alpha(q))$ sur de petites échelles de temps (minute) et plus de variétés de valeurs de $f(\alpha(q))$ sur des échelles de temps plus grandes selon différentes valeurs de q avec une forte concentration de $f(\alpha(q))$ lorsque $q \rightarrow -\infty$.

Nous constatons que la dynamique Standard & Poor500 varie fortement selon les différentes échelles de temps, présentant des statistiques différentes et plus de complexité au sens de Shannon à mesure que les échelles de temps deviennent plus grandes. Enfin, nous concluons que la valeur de l'entropie plus faible sur des échelles de temps plus petites doit être comprise par le fait que le dynamisme des rendements logarithmiques exhibe une information plus prévisible avec moins d'incertitude, donc exhibe moins d'information spécifique au sens de Shannon.

Nous proposons dans le chapitre 3 de modéliser ce phénomène de complexité sous-jacent aux séries financières via la mécanique statistique et la super-statistique. Nous montrerons que nous pouvons considérer l'origine de cette complexité comme étant l'émergence d'une volatilité fluctuante à l'échelle macroscopique du système, définie selon une superposition de statistiques inhomogènes à l'échelle locale.

Chapitre 3 : Modélisation de la complexité : approche super statistique

3.1 Introduction

Dans ce Chapitre, nous nous intéressons au processus de modélisation de la dynamique complexe de certains systèmes telle que les séries financières. Nous introduisons un formalisme généralisé de la mécanique statistique concernant les systèmes dont le comportement se caractérise par une superposition de plusieurs statistiques. Ce formalisme fut principalement développé par Beck, C [14, 13, 112] qui propose dans ses travaux le terme de super-statistique pour décrire ce type de phénomènes. A titre d'exemple, Beck, C propose de considérer le comportement d'une particule Brownienne se déplaçant dans un environnement changeant. Nous observons alors deux types de dynamiques. Une dynamique relativement rapide issue de la vitesse de la particule brownienne et une dynamique relativement lente correspondant au changement de température, à l'origine du changement d'environnement. Ces deux effets sont associés selon des échelles de temps différentes et il en résulte ainsi une superposition de deux statistiques. L'approche essentielle de ce formalisme est que la séparation d'échelle temporelle entre ces deux effets est suffisamment importante pour que le macro-état du système ait assez de temps pour se relaxer vers un état d'équilibre et y rester pendant pour un certain temps.

Les super-statistiques ont de nombreuses applications pratiques dans la modélisation de systèmes complexes, notamment dans la modélisation de turbulences hydrodynamiques [15, 97, 111], la dynamique des prix boursiers [87, 109, 23, 6], la théorie des matrices aléatoires [1, 2, 3], l'étude de la fluctuation de la vitesse du vent [99, 73], l'étude des fluctuations hydro-climatiques [93], l'étude statistique des retards de départ de trains [20], ou encore l'étude statistique de survie des patients atteints de cellules cancéreuses [27].

L'objectif principal de ce chapitre est de proposer une méthode alternative aux modèles standards de modélisations du processus de volatilité des séries financières, en étudiant le processus physique sous-jacent à ces séries. Nous considérons les séries de rendements logarithmiques d'un actif quelconque comme étant un système dynamique non-linéaire dont le caractère complexe émergent est le fruit d'un environnement fluctuant, à savoir une variance non constante dont la dynamique change selon l'échelle spatio-temporelle d'étude. Nous considérons ainsi deux statistiques, une relativement rapide qui correspond à la vélocité des prix logarithmiques et qui tend à relaxer vers une distribution Gaussienne, et une deuxième relativement lente correspondant à la distribution de l'ensemble des variances locales de la série. La distribution stationnaire de ce système super-statistique obtenu en moyennant l'ensemble des variances locales peut, selon l'échelle temporelle d'étude, faire émerger un comportement non-Gaussien se caractérisant par une Loi de puissance et des queues de distributions grasses.

Ce chapitre s'organise autour de 4 sections. Dans un premier temps, nous proposons d'introduire au travers de la mécanique statistique de Maxwell-Boltzmann le concept de super-statistique. Nous montrons également que les principales familles de super-statistiques sont en réalité des cas particuliers de statistiques plus générales.

Puis, en reprenant les hypothèses de la théorie des mouvements boursiers d'Osborne -et au travers d'une analogie avec la mécanique statistique, nous proposons deux expériences numériques simulant la distribution des opportunités de profits d'agents financiers selon que nous considérons les rendements logarithmiques comme étant un système conservatif ou dissipatif. Ces simulations permettent de mettre en avant le concept de superposition de statistiques, de variance non constante, de modifier la théorie d'Osborne et d'introduire la notion de volatilité stochastique. Nous montrons notamment qu'il est possible par inférence Bayésienne de postuler que la fonction inverse gamma est la probabilité conjuguée aux rendements logarithmiques permettant de modéliser le processus fluctuant de la variance.

Dans une troisième section, nous proposons de modéliser par inférence Bayésienne, pour des échelles de temps intra-journalières différentes, le processus de la variance au moyen de la statistique inverse gamma et log-normal. Nous élargirons l'étude en modélisant la distribution des rendements logarithmiques au travers du produit de la probabilité de vraisemblance des rendements avec la distribution de la variance. La conclusion principale de cette étude est que ces deux super-statistiques sont significatives dans la modélisation de données sur des échelles de temps hautes fréquences, mais perdent en significativité à mesure que l'échelle de temps s'élargit.

Par conséquent, nous proposons, dans une quatrième section, de prouver l'existence d'une transition de statistiques de la distribution du paramètre de variance. Finalement, nous résumerons l'ensemble de nos résultats.

3.2 Super-statistiques : principes fondateurs

3.2.1 Maxwell-Boltzmann statistique

Toutes les molécules d'air qui nous entourent ne se déplacent pas à la même vitesse. Certaines molécules d'air se déplacent plus rapidement que d'autres. Certaines se déplacent à des vitesses modérées et certaines molécules restent presque immobiles. Ainsi, comment répondre à la question suivante : Quelle est la vitesse d'une molécule d'air dans un gaz, puisqu'elle pourrait avoir n'importe laquelle d'un grand nombre de vitesses possibles ? Si on imagine un gaz idéal qui est constitué d'un nombre infiniment grand de molécules interagissant les unes avec les autres, du fait du degré de liberté non négligeable, il est intuitivement compliqué d'étudier le comportement de chaque molécule au travers de leur position et vitesse à chaque instant.

Par conséquent, au lieu de s'intéresser à la vitesse d'une molécule particulière d'un gaz quelconque, la mécanique statistique développée par James Maxwell et Ludwig Boltzmann au milieu du XIX^{ème} siècle propose d'étudier la distribution de la vitesse d'un gaz selon une certaine température. Ils proposent une théorie probabiliste permettant de représenter la distribution des vitesses de l'ensemble des molécules d'un

gaz parfait. Cette théorie fut présentée dans un des articles publiés par Maxwell en 1867 [83] dans lequel il donne le cadre théorique aux calculs de vitesses, du libre parcours moyen et du nombre de collisions des molécules d'un gaz à une température donnée.

Cette distribution de vitesses est appelé la distribution de Maxwell-Boltzmann et est souvent représentée pour un gaz parfait à l'aide du graphique suivant :

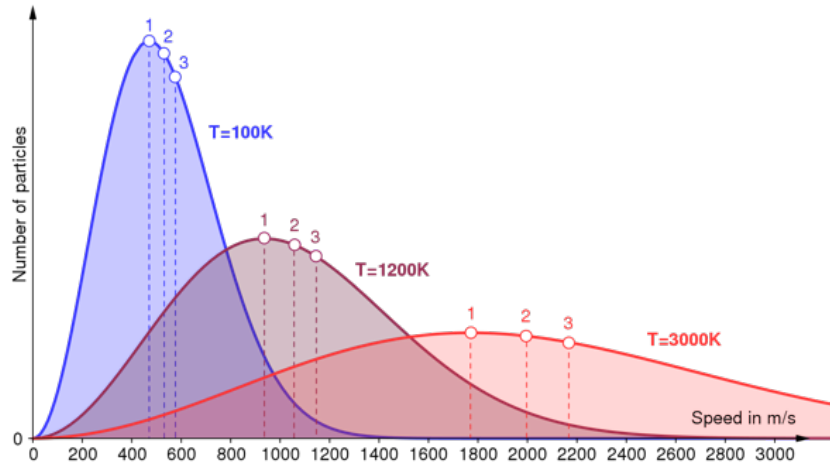


Figure 3. 1 Distribution de Maxwell-Boltzmann pour un gaz idéal selon différentes températures données (<https://commons.wikimedia.org/wiki/File:Maxwell-Boltzmann-Distribution-simple-axis-labels.svg>).

L'axe des ordonnées du graphique de profils de vitesses d'un gaz parfait est considéré comme donnant le nombre de molécules par unité de vitesse. On note que ce graphique n'est pas symétrique ; il présente une queue de distribution pour de larges vitesses et la vitesse ne peut qu'être une valeur strictement positive. Si la température du gaz augmente, le pic de la distribution se déplacera vers la droite car la vitesse moyenne moléculaire augmente et inversement pour un gaz qui se refroidit.

Considérons comme précédemment un gaz idéal constitué d'un nombre de molécules suffisamment important pour qu'elles rentrent en collision. Nous définissons le macro-état Ω du système comme étant caractérisé par trois composantes, à savoir le nombre de molécules le constituant, le volume du système et l'énergie du système. Un macro-état peut être caractérisé par différents micro-états selon la configuration de l'ensemble de ses constituants. Nous définissons un micro-état par ω et représente une

représentation particulière de Ω . On rappelle que tout système peut être vu comme un sous-ensemble d'un système plus large (sauf exception de l'univers qui ne peut être un sous-système). Ainsi, s'intéresser au comportement local de certaines molécules revient à définir arbitrairement un sous-système. Et si un système n'échange aucune forme d'énergie ou de matière avec son environnement extérieur, alors il est dit conservatif. A l'inverse, il sera considéré comme dissipatif (Figure 3.2).

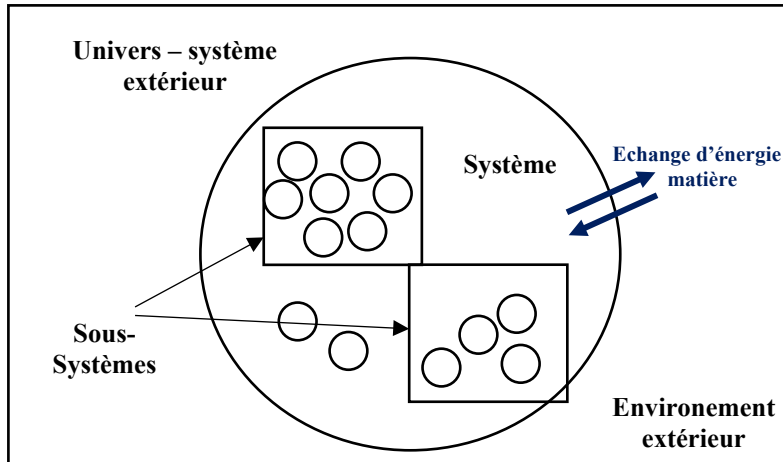


Figure 3. 2 Représentation schématique d'un système comme ensemble physique constitué d'un ensemble de sous-systèmes.

Le nombre total de molécules du système est défini par n comme une grandeur fixe. On définit le volume total du système par V comme fixe et l'énergie totale du système par U comme étant également une valeur fixe. Chaque molécule du système se caractérise par un niveau d'énergie donnée par $\varepsilon_0, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_r$. Le nombre de molécules indexées à chaque niveau d'énergie est variable et est défini par $n_0, n_1, n_2, \dots, n_r$ et $\sum_{i=0}^r n_i = n$ et $\sum_{i=0}^r n_i * \varepsilon_i = U$. La question de savoir comment l'énergie U est répartie parmi les molécules et comment ces molécules sont distribuées dans le système, c'est-à-dire combien de micro-états peuvent exister dans ce système, représentent deux questions fondamentales à l'élaboration de la théorie de Maxwell-Boltzmann. En effet, la répartition et le déplacement des molécules sont liés aux échanges d'énergies du fait de leurs interactions. Et, dans le cas d'un gaz, c'est le phénomène de thermalisation, à savoir l'échange d'énergie cinétique dans toutes les directions issues des interactions entre molécules, qui est à l'origine du comportement

dynamique du système. Et plus la température est élevée, plus le comportement du système est dynamique avec une augmentation de la vitesse moyenne des molécules (Figure 3.3). Ainsi, selon la température, nous serons en présence de différents types de micro-états ω , c'est pourquoi nous considérons la température comme source de désordre d'un système et le nombre de désordres d'un système est maximal, la probabilité d'obtenir un certain micro-état étant maximisée.

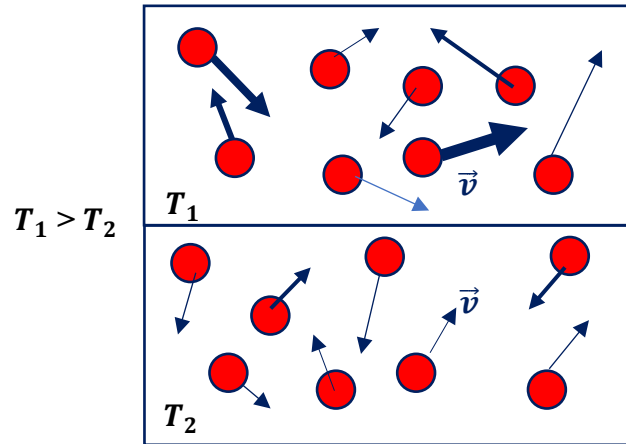


Figure 3. 3 Représentation du comportement de molécules d'un gaz selon l'activité thermique du système. La température est liée à l'énergie cinétique moyenne des composants du système.
 $\langle E_c \rangle = \frac{1}{2} m \langle v^2 \rangle = \frac{3}{2} k_b T$, avec E_c l'énergie cinétique, v la vitesse, T la température et k_b la constante de Boltzmann.

Le nombre de combinaisons possibles pour atteindre un certain micro-état est donnée par la formule :

$$\omega = \frac{n!}{n_0! n_1! n_2! \dots n_r!} \quad (3.1).$$

Ainsi, on cherche à déterminer la combinaison de $n_0, n_1, n_2, \dots, n_r$ qui permet de maximiser ω . Cette combinaison doit aboutir au micro-état le plus probable et sera considérée comme l'état d'équilibre. Cet équilibre est trouvé en résolvant l'équation suivante :

$$\ln(\omega) = \frac{\ln(n!)}{\prod_{i=0}^r n_i!} \quad (3.2)$$

$$\ln(\omega) = \ln(n!) - \sum_{i=0}^r \ln(n_i!) \quad (3.3)$$

D'après la méthode d'approximation de Sterling [...] on a :

$$\ln(x!) = x \ln(x) - x + \mathcal{O}(\ln(x)) \quad (3.4)$$

Ainsi, la relation (3.3) peut être approximée comme suit :

$$\ln(\omega) = n \ln(n) - n - \sum_{i=0}^{i=r} [n_i \ln(n_i) - n_i] \quad (3.5)$$

En dérivant l'équation (3.5), on obtient :

$$\partial \ln(\omega) = - \sum_{i=0}^{i=r} \left[\partial n_i \ln(n_i) + n_i * \frac{1}{n_i} - \partial n_i \right] = 0 \quad (3.6)$$

Équation (3.6) implique que :

$$\sum_{i=0}^{i=r} \partial n_i \ln(n_i) = 0 \quad (3.7)$$

Et comme $\sum_{i=0}^{i=r} n_i = \text{constant}$ (3.8) par hypothèse, par implication on obtient :

$$\sum_{i=0}^{i=r} \partial n_i = 0. \quad (3.9)$$

Nous avons défini ci-dessus l'hypothèse que l'énergie totale du système était constante telle que $\sum_{i=0}^r n_i * \varepsilon_i = U = \text{constant}$ (3.10). D'où, en différentiant l'équation précédente par rapport à ε_i , on obtient :

$$\sum_{i=0}^{i=r} \varepsilon_i \partial n_i = U = 0. \quad (3.11)$$

Afin de maximiser l'équation (3.7) soumise aux contraintes (3.9) et (3.11), nous utilisons la méthode des multiplicateurs de Lagrange. Cette méthode permet de trouver les points maximum et minimum d'une fonction dérivable sous contraintes. En

multipliant l'équation (3.9) par un paramètre α et l'équation (3.11) par un paramètre β , puis en additionnant les équations (3.7), (3.9) et (3.11), on obtient :

$$\sum_{i=0}^{i=r} [\partial \ln(n_i) + \alpha + \beta \varepsilon_i] \partial n_i = 0 \quad (3.12)$$

Selon l'équation (3.12), on en déduit que :

$$\partial \ln(n_i) + \alpha + \beta \varepsilon_i = 0 \quad (3.13)$$

$$n_i = e^{-\alpha} * e^{-\beta \varepsilon_i} \quad (3.14)$$

L'équation (3.14) représente le nombre de molécule au nième niveau. Le seul paramètre variable est le niveau d'énergie au nième niveau à savoir ε_i . Ainsi, le nombre total de molécules du système est :

$$\sum_{i=0}^{i=r} n_i = e^{-\alpha} * \sum_{i=0}^{i=r} e^{-\beta \varepsilon_i} = n \quad (3.15)$$

Par simplification on obtient :

$$e^{-\alpha} = \frac{n}{\sum_{i=0}^{i=r} e^{-\beta \varepsilon_i}} = \frac{n}{P} \quad (3.16),$$

avec P qui représente en mécanique statistique la fonction de partition. Cette fonction est une grandeur fondamentale qui regroupe les propriétés statistiques d'un système à l'équilibre thermodynamique.

En substituant (3.16) dans (3.14), nous obtenons finalement la distribution de Maxwell-Boltzmann :

$$n_i = \frac{n}{P} * e^{-\beta \varepsilon_i} \quad (3.17).$$

Ici, le paramètre $\beta = \frac{-1}{k_b T}$, où k_b est la constante de Boltzmann et T est la température.

En substituant la valeur de β dans l'équation (3.17), on obtient :

$$n_i = \frac{n}{P} * e^{\frac{\varepsilon_i}{k_b T}} \quad (3.18).$$

L'expression (3.18) permet de déterminer le nombre de particules du micro-état le plus probable. Cette expression est également connue comme l'expression de la statistique de Maxwell-Boltzmann.

Nous notons que la probabilité de trouver une molécule au nième niveau est proportionnelle à n_i telle que $p_i \sim n_i$. Ainsi, en considérant toujours U comme l'énergie moyenne du système de l'ensemble des énergies ε_i de différents niveaux, alors, en modifiant l'équation (3.10), on peut écrire :

$$U_{loc}(\beta) = \sum p_i * \varepsilon_i \quad (3.19),$$

avec $U_{loc}(\beta)$ l'énergie moyenne locale interne d'une ou d'un groupe de cellules au nième niveau avec $\beta = 1/T$, l'inverse de la température. L'entropie est donnée sous la forme de Boltzmann-Gibbs-Shannon et prend en compte l'énergie locale ε_i :

$$S = -k_b \sum_i p_i \log(p_i) \quad (3.20)$$

$$S = -k_b \sum_i \frac{n}{p} * e^{-\beta \varepsilon_i} * \left(-\log\left(\frac{n}{p}\right) - \beta \varepsilon_i \right) \quad (3.21)$$

$$S = \beta U + \log\left(\frac{n}{p}\right) \quad (3.22)$$

Finalement, la formule de Boltzmann est donnée par la relation suivante :

$$S = k_b \log(\omega) \quad (3.23)$$

La formule de Boltzmann décrit la relation qu'il existe entre l'entropie et la probabilité $p_i = 1/\omega$ d'avoir une configuration de micro-états donnée pour un niveau d'énergie connu. L'entropie est donc maximale lorsque le micro-état atteint est celui qui se caractérise par une configuration de molécules la plus probable. Le système atteint ainsi son équilibre lorsque son entropie est maximale.

3.2.2 Généralisation de la statistique de Boltzmann

Le terme de super-statistique a été énoncé pour la première fois en 2003 par Beck, C et Cohen [13], issu des travaux précédents publiés par Beck, C et Wilk & Wlodarczyk [120] dans lesquels ils donnent une nouvelle interprétation physique du comportement de certains systèmes complexes, en montrant notamment que les systèmes en non-équilibre subissant des fluctuations de températures peuvent être décrits de manière efficace par une généralisation de la statistique de Boltzmann. Ils illustrent ce concept en considérant l'équation (3.17) :

$$n_i = \frac{n}{P} * e^{-\beta \varepsilon_i}$$

$$P \sum_i \frac{n_i}{n} = \sum_i e^{-\beta \varepsilon_i} \quad (3.24)$$

$$P = \sum_i e^{-\beta \varepsilon_i} \quad (3.25)$$

$$P \sim \int_0^\infty f(\beta) e^{-\beta E} d\beta \quad (3.26)$$

où P est la fonction de partition et E l'énergie moyenne.

$$\int_0^\infty f(\beta) e^{-\beta E} d\beta = \frac{1}{(1+(q-1)\beta_0 E)^{1/(q-1)}} \quad (3.27),$$

Où :

$$f(\beta) = \frac{1}{\Gamma\left(\frac{1}{q-1}\right)} \left\{ \frac{1}{(q-1)\beta_0} \right\}^{1/(q-1)} \beta^{\left(\frac{1}{q-1}-1\right)} \exp\left\{ \frac{1}{(q-1)\beta_0} \right\} \quad (3.28),$$

avec $f(\beta)$ est la probabilité de distribution gamma (Γ), β_0 et q sont les paramètres de la distribution, avec $q > 1$ et β_0 la valeur moyenne du paramètre fluctuant β . Nous constatons que la moyenne du facteur ordinaire de Boltzmann $e^{-\beta E}$, avec le paramètre β distribué selon une distribution gamma, permet d'obtenir un facteur de Boltzmann de la forme $q - exponentielle$ donné par la partie droite de l'équation (3.27), facteur qui représente également la fonction de partition. Ainsi, en généralisant la statique de Boltzmann, Beck et Willk & Wlodarczyk, montrent que la distribution gamma est celle qui permet de manière effective d'un point de vue dynamique de modéliser le comportement du paramètre β .

L'approche plus pratique de l'utilisation des super-statistiques fut plus développée par Beck, C et Cohen. Ils définissent un système complexe comme un système en non-équilibre et spatio-temporel inhomogène. Un tel système peut être représenté sous une forme de « coarse-grained », c'est-à-dire qu'il peut être divisé en plusieurs cellules se caractérisant toutes par la présence d'un paramètre fluctuant β . Dans le cas de la statistique de Boltzmann pour l'étude d'un gaz, ce paramètre est l'inverse de la température. De manière générale, ce paramètre peut être un facteur d'une équation stochastique, le paramètre de volatilité en finance ou encore une variance locale extraite de certaines séries temporelles quelconques notamment des séries financières. Localement, chaque cellule définie par une constante β est supposée atteindre l'équilibre selon un temps de relaxation relativement court.

Ainsi, en considérant les résultats de Beck, C et Willk & Wlodarczyk mis en avant par l'équation (3.28), si nous nous intéressons au comportement de la température ou d'une variance locale, nous nous intéresserons au paramètre $1/\beta$, c'est-à-dire $f(1/\beta)$.

Finalement, le terme de super-statistique est une abréviation de la notion de superposition d'une ou plusieurs statistiques.

Beck, C [16] propose également de construire mathématiquement le comportement dynamique des super-statistiques au travers des équations de Langevin composées de paramètres fluctuants selon les échelles de temps. Ces équations de Langevin à l'échelle locale décrivent le comportement mésoscopique du système. L'exemple le plus simple serait une équation locale de Langevin que Beck, C décrit comme suit :

$$\dot{v} = -\gamma v + \sigma L(t) \quad (3.29),$$

avec γ et σ des paramètres fluctuant relativement lentement et $L(t)$ un bruit Gaussien. L'équation (3.29) décrit l'évolution de la vitesse v d'une particule Brownienne qui se

déplace au travers de cellules spatiales caractérisées par un paramètre local β défini dans chaque cellule par $\gamma/2\sigma^2$ représentant ainsi une situation de non-équilibre. La Figure 3.4 en donne une représentation simple :

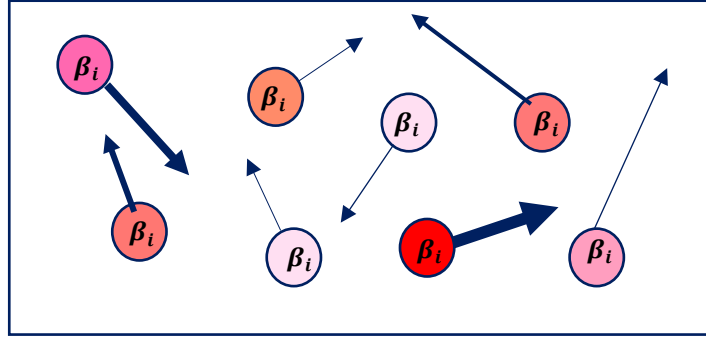


Figure 3. 4 Représentation d'une particule Brownienne dans un environnement spatio-temporel inhomogène du fait d'une température inverse β_i différente selon les régions du système (Figure inspirée de l'article [14]).

Si une distribution de probabilité $f(\beta)$ de la température inverse β pour différentes cellules est donnée par équation (3.28), alors la distribution de probabilité de la vitesse v conditionnée au paramètre β est une distribution Gaussienne telle que :

$$p(v|\beta) = \sqrt{\frac{\beta}{2\pi}} e^{-(1/2)\beta v^2} \sim e^{-(1/2)\beta v^2} \quad (3.30).$$

La probabilité jointe de v, β est donnée par :

$$p(v, \beta) = f(\beta)p(v|\beta) \quad (3.31)$$

Et la probabilité marginale $p(v)$ est définie comme suit :

$$p(v) = \int_0^\infty f(\beta)p(v|\beta)d\beta \quad (3.32)$$

Par conséquent, l'équation (3.32) est considérée comme étant une super-statistique car c'est une superposition d'une statistique $f(\beta)$ et d'une distribution

Gaussienne. Autrement dit, Beck, C caractérise un système complexe par une superposition de plusieurs statistiques sur différentes échelles de temps :

- une statistique issue de la mécanique statistique traditionnelle modélisée à l'échelle mésoscopique par l'équation de Langevin,
- et une seconde statistique correspondant à un paramètre qui fluctue de manière relativement lente qui, pour le cas d'un gaz, correspond à l'inverse de la température et, dans le cas des séries financières, correspond à l'inverse de la variance.

Les distributions stationnaires des systèmes super-statiques, obtenus en faisant la moyenne de l'ensemble des β du système, présentent généralement un comportement non-Gaussien caractérisé par des queues de distributions grasses qui peuvent être en forme de Loi de puissance, ou selon une Loi exponentielle étirées. D'autres fonctions sont aussi à considérer et font l'objet de la section suivante.

3.2.3 Description des familles fondamentales de super-statistiques

La distribution des familles de super-statistique est déterminée par la structure dynamique observable à l'échelle "macro-état" du système complexe étudié. Les différentes formes de distributions de $f(\beta)$ peuvent être déterminées pour des contraintes données par le principe de maximisation de l'entropie du système et font référence également aux travaux de Van der Straeten & Beck [17, 109]. Puis, dans un autre article, et se basant sur les travaux précédents, Beck, C et Cohen [15] cherchent à déterminer quel type de fonction $f(\beta)$ est généralement observée pour des données empiriques, données issues du comportement d'un système complexe. Ils montrent que, comme β est un paramètre strictement positif déterminé par une dynamique spatio-temporelle inhomogène, une distribution Gaussienne ne peut être considérée. A l'inverse, ils montrent que trois super-statistiques dans l'application pratique de modélisation de systèmes complexes sont fréquentes, à savoir la distribution gamma (Ga), la distribution inverse gamma (IGa) et la distribution log-normale (LN).

Nous montrons ci-dessous que ces trois statistiques sont des cas particuliers d'une statistique issue d'une distribution généralisée.

- Dans le cas où la distribution $f(\beta)$ est donnée par la fonction gamma, alors nous montrons que c'est un cas particulier de la distribution gamma généralisée (GGa) telle que :

$$GGa(x; \alpha, \beta, \gamma) = \frac{\gamma}{\beta\Gamma(\alpha)} e^{-\left(\frac{x}{\beta}\right)^\gamma} \left(\frac{x}{\beta}\right)^{-1+\alpha\gamma} \quad (3.33),$$

avec pour $x > 0$ ou égal à 0. Les paramètres $\alpha, \beta, \gamma > 0$. Et, pour $\gamma = 1$, la fonction gamma généralisée devient une fonction gamma définie comme suit :

$$Ga(x; \alpha, \beta) = \frac{1}{\beta\Gamma(\alpha)} e^{-\left(\frac{x}{\beta}\right)} \left(\frac{x}{\beta}\right)^{-1+\alpha} \quad (3.34)$$

$$Ga(x; \alpha, \beta) = \frac{\beta^{-1} \beta^{1-\alpha}}{\Gamma(\alpha)} e^{-\left(\frac{x}{\beta}\right)} x^{-1+\alpha} \quad (3.35)$$

$$Ga(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \left(\frac{1}{\beta}\right)^\alpha e^{-\left(\frac{x}{\beta}\right)} x^{\alpha-1} \quad (3.36)$$

Ainsi, en substituant dans l'équation (3.36) le paramètre β à x et le paramètre β_0 à β , on obtient une distribution de β selon les paramètres α et β_0 comme suit :

$$f(\beta; \alpha, \beta_0) = \frac{1}{\Gamma(\alpha)} \left(\frac{1}{\beta_0}\right)^\alpha e^{-\left(\frac{\beta}{\beta_0}\right)} \beta^{\alpha-1} \quad (3.37)$$

- Dans le cas le cas où la distribution $f(\beta)$ est donnée par la fonction inverse gamma : alors nous montrons que c'est un cas particulier de la distribution inverse gamma généralisée (GIGa) telle que :

$$GIGa(x; \alpha, \beta, \gamma) = \frac{\gamma}{\beta\Gamma(\alpha)} e^{-\left(\frac{x}{\beta}\right)^{-\gamma}} \left(\frac{x}{\beta}\right)^{-1-\alpha\gamma} \quad (3.38),$$

pour $x > 0$ ou égal à 0. Les paramètres $\alpha, \beta, \gamma > 0$. Et, pour $\gamma = 1$, la fonction inverse gamma généralisée devient une fonction inverse gamma définie comme suit :

$$IGa(x; \alpha, \beta) = \frac{1}{\beta\Gamma(\alpha)} e^{-\left(\frac{\beta}{x}\right)} \left(\frac{\beta}{x}\right)^{1+\alpha} \quad (3.39)$$

$$IGa(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \beta^\alpha x^{-1-\alpha} e^{-\left(\frac{\beta}{x}\right)} \quad (3.40)$$

Ainsi, en substituant dans l'équation (3.36) le paramètre β à x et le paramètre β_0 à β , on obtient une distribution de β selon les paramètres α et β_0 comme suit :

$$f(\beta; \alpha, \beta_0) = \frac{1}{\Gamma(\alpha)} \beta_0^\alpha \beta^{-1-\alpha} e^{-\left(\frac{\beta_0}{\beta}\right)} \quad (3.41).$$

Nous ajoutons que, dans la distribution inverse gamma, la fonction $f(\beta)$ a une queue de distribution en loi de puissance du type $\beta^{-1-\alpha}$ pour $\beta \gg \beta_0$.

- Dans le cas où la distribution $f(\beta)$ est donnée par la fonction log-normale : En s'appuyant sur les travaux de Tao Ma et R.A Serota [76], nous sommes en mesure de montrer que cette dernière est le cas limite de la distribution inverse gamma généralisée pour $\alpha \rightarrow \infty$.

L'équation (3.41) peut être écrite sous la forme généralisée :

$$f(\beta; \alpha, \beta_0, \gamma) = \frac{\gamma}{\beta \Gamma(\alpha)} \exp \left[\alpha \left(\frac{\beta_0}{\beta} \right)^\gamma - \left(\frac{\beta_0}{\beta} \right)^\gamma \right] \quad (3.42)$$

On définit les différents paramètres qui caractérisent une distribution log-normale comme suit :

$$\mu = \ln(\beta_0) - \frac{1}{\gamma} \ln \left(\frac{1}{\gamma^2} \right) \quad (3.43)$$

$$s = \frac{1}{\gamma \sqrt{\alpha}} \quad (3.44)$$

$$\delta = \frac{1}{\sqrt{\alpha}} \quad (3.45)$$

Avec $s > 0$ et $\delta > 0$ on obtient :

$$\alpha = \frac{1}{\delta^2} \quad (3.46)$$

$$\beta_0 = e^\mu \delta^{\frac{-2s}{\delta}} \quad (3.47)$$

$$\gamma = \frac{\delta}{s} \quad (3.48)$$

Et :

$$\left(\frac{\beta_0}{\beta} \right)^\gamma = \delta^{-2} e^{\frac{-\delta}{s} (\ln(\beta) - \mu)} \quad (3.49)$$

$$\ln \left(\frac{\beta_0}{\beta} \right)^\gamma = \frac{-\delta}{s} * (\ln(x) - \mu) + \ln(\delta^{-2}) \quad (3.50)$$

Tao et Serota proposent de simplifier l'équation (3.49) par une expansion de Taylor ; avec $\frac{\delta}{s} \rightarrow 0^+$, on obtient :

$$\alpha \left(\frac{\beta_0}{\beta}\right)^\gamma - \left(\frac{\beta_0}{\beta}\right)^\gamma \sim \frac{\ln(\delta^{-2})-1}{\delta^2} - \frac{(\ln(\beta)-\mu)^2}{2s^2} \quad (3.51).$$

D'où, en substituant l'équation (3.51) dans (3.42), on a :

$$f(\beta; \alpha, \beta_0, \gamma) = \frac{\gamma}{\beta\Gamma(\alpha)} \exp\left[\frac{\ln(\delta^{-2})-1}{\delta^2}\right] \exp\left[-\frac{(\ln(\beta)-\mu)^2}{2s^2}\right] \quad (3.52).$$

Tao et Serota utilisent une approximation du membre de l'équation (3.52), à savoir $\frac{\gamma}{\Gamma(\alpha)} \exp\left[\frac{\ln(\delta^{-2})-1}{\delta^2}\right]$, au travers de l'utilisation de la formule asymptotique de Sterling en laissant $\delta^{-2} = \alpha \rightarrow \infty$ et, par conséquent, $\delta \rightarrow 0$.

On obtient finalement que :

$$\frac{\gamma}{\Gamma(\alpha)} \exp\left[\frac{\ln(\delta^{-2})-1}{\delta^2}\right] \sim \frac{1}{s\sqrt{2\pi}} \quad (3.53)$$

En substituant l'équation (3.53) dans (4.52) on obtient :

$$f(\beta; \mu, s, \gamma) = \frac{\gamma}{\beta s\sqrt{2\pi}} \exp\left[-\frac{(\ln(\beta)-\mu)^2}{2s^2}\right] \quad (3.54)$$

Finalement, pour $\delta \rightarrow 0$ et $\gamma = 1$, on obtient une distribution log-normale définie selon l'équation suivante :

$$f(\beta; \mu, s) = \frac{1}{s\beta\sqrt{2\pi}} \exp\left[-\frac{(\ln(\beta)-\mu)^2}{2s^2}\right] \quad (3.55) \text{ avec } \beta, s > 0.$$

3.3 Superposition de statistiques des séries financières : expériences numériques¹⁰

Cette section examine la théorie du mouvement brownien d'Osborne pour la dynamique des prix boursiers [88]. L'hypothèse d'Osborne est modifiée pour tenir compte de l'asymétrie d'opportunité de profit entre agents financiers. Au travers de simulations numériques, nous étudions les dynamiques sous-jacentes à l'évolution des cours boursiers, en proposant deux expériences selon lesquels le cours boursier est supposé comme étant soit un système conservatif, soit système dissipatif.

¹⁰ Cette section 3.3 fait l'objet d'un article publié dans le journal « Economics Bulletin »

Dans un système conservatif, nous montrerons que la dynamique converge rapidement vers un équilibre global, qui est la probabilité marginale de profit dans un jeu non biaisé (c'est-à-dire une probabilité de 0,5), confirmant les implications de la théorie d'Osborne. Cependant, des asymétries existent en raison des hétérogénéités d'opportunité entre les agents.

Dans un système dissipatif, nous montrerons que la dynamique peut être décrite par une superposition de distributions en équilibre local à variance constante satisfaisant localement les implications de la théorie d'Osborne. En revanche, globalement, la dynamique n'atteint pas l'équilibre, et sa distribution présente des asymétries d'opportunités et se caractérise par une variance fluctuante.

Dans les deux expériences, nous montrerons que la loi de t-Student est celle qui décrit de manière significative l'évolution de l'opportunité de profit du système.

Enfin, nous étendrons notre travail et fournirons des preuves empiriques que le dynamisme de la variance de l'indice Standard&Poors 500 subit une transition statistique d'un horizon temporel de hautes fréquences à un horizon de basses fréquences. Sous haute fréquence (c'est-à-dire intra-journalière), les distributions log-normale et inverse gamma expliquent de manière équivalente et significatif le dynamisme de la variance du système. En revanche, sur des fréquences temporelles basses (c'est-à-dire « hebdomadaire » et « mensuelle »), la distribution gamma est préféré.

3.3.1 Théorie d'Osborne concernant le mouvement Brownien des cours boursiers

La théorie du mouvement brownien d'Osborne de la dynamique des prix des cours boursiers repose sur l'hypothèse d'une parfaite symétrie d'opportunité de profits entre les agents financiers et implique que, sur des intervalles de temps non chevauchant, cette symétrie est conservée et la dynamique des prix des cours boursiers constitue une marche aléatoire définie par une fonction de probabilité commune à variance constante.

Pour parvenir à cette conclusion, Osborne fait plusieurs hypothèses :

- Le logarithme des prix est le rapport approprié pour mesurer la variation des prix. Si $p(t)$ représente le prix d'une action à l'instant t , alors $x_t = \ln(p_{t+\tau}) - \ln(p_t)$ est la variation du log du prix de l'instant t à $t + \tau$.
- Les prix peuvent être considérés comme un ensemble de molécules en équilibre en mécanique statistique. Les distributions de probabilité de l'état stationnaire sont déterminées par la condition de probabilité maximale et la distribution d'équilibre est donnée par :

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \quad (3.56)$$

où σ^2 est la variance de x et où la moyenne est considérée comme nulle. Par conséquent, la distribution est la même que celle d'une particule en mouvement brownien.

- La variance σ^2 des variations de prix est constante sur l'intervalle de temps unitaire.
- La condition la plus probable, sous laquelle une transaction financière sur le marché boursier est possible, est obtenue lorsque les opportunités de profits attendues du vendeur et de l'acheteur pour chaque pari sont maximisées et, à long terme, sont égales. Autrement dit, à long terme -et selon Osborne-, la somme des bénéfices attendus de l'acheteur et du vendeur est nulle et le temps nécessaire pour atteindre cet équilibre est déterminé par une variance finie qui augmente en fonction de la racine carrée du nombre de transactions. Dès lors, une transaction devient possible lorsque l'estimation de profits de l'acheteur et du vendeur sont équivalentes et opposées, de sorte que : $E_{acheteur} \cdot \Delta \log(P) + E_{vendeur} \cdot \Delta \log(P) = 0$, où P est le prix d'une action quelconque, et E est l'espérance moyenne de profits des agents. Ainsi, si E est égal à une probabilité marginale de 0,5, alors l'hypothèse d'Osborne est vérifiée.

Cette « symétrie d'opportunité » en tant qu'hypothèse a été examinée et remise en question dans plusieurs études et par plusieurs auteurs, dont Osborne lui-même [88],

Mandelbrot [79] et Fama ¹¹ et, plus récemment, Logfren, Persson et Weibull [118] qui mettent l'accent sur la preuve que l'asymétrie de l'information émergeant des marchés financiers est une caractéristique et une condition nécessaire aux interactions entre agents. Une conséquence de la théorie d'Osborne est que les agents financiers sont homogènes, partagent la même information commune et, par conséquent, prennent la décision de parier ou non sur le marché selon un processus de décision identique.

Dans la suite de cette section, nous gardons une des hypothèses d'Osborne. Nous considérons que les agents financiers prennent une décision si, et seulement si, ces derniers maximisent leur opportunité de profits compte tenu des informations disponibles. Nous définissons l'information dont disposent les agents comme étant leur probabilité de profits conditionnée par la série historique de résultats sur leurs paris précédents. En d'autres termes, nous considérons que chaque agent met à jour son opportunité de profits espérés par inférence bayésienne en prenant chaque nouveau résultat comme une information nouvelle et supplémentaire par rapport à son état actuel. Nous considérons le marché comme un système constitué de groupes locaux d'agents homogènes partageant des informations identiques (c'est-à-dire la même probabilité conditionnelle de profits). En revanche, l'ensemble des groupes du système constitue un ensemble hétérogène (c'est-à-dire des groupes d'agents avec différentes probabilités conditionnelles de profits espérés). C'est ce qu'on appelle un système de type « coarse-grained » : un système constitué d'un maillage de sous-ensembles locaux inhomogènes.

Cette représentation du système en « coarse-grained » sera notre seule hypothèse dans la suite de l'expérience. Nous ne portons aucun préjugé sur la distribution sous-jacente des rendements logarithmiques et cherchons à comprendre sa dynamique au travers de deux expérimentations :

- Considérer la dynamique des prix comme un système conservatif. Le système est isolé sans aucune interaction avec son environnement. Le système commence et finit la simulation avec le même nombre d'agents financiers.

¹¹ Fama E. F. Mandelbrot and the stable Paretian hypothesis. *Journal of Business*, 36:420–429, 1963.

- Considérer la dynamique des prix comme un système dissipatif. Le système interagit avec son environnement, permet à de nouveaux agents d'entrer sur le marché et aux agents d'en sortir s'ils considèrent avoir une opportunité de profits trop faible ou s'ils font faillite.

Dans la section suivante, nous mettons en avant les résultats des deux expériences numériques. Nous montrons des preuves statistiques d'asymétrie d'opportunité de profits à long terme entre agents. Nous montrons que ces asymétries sont la conséquence d'une statistique sous-jacente à la dynamique des cours boursiers et qui vient contredire l'hypothèse d'Osborne de variance constante.

3.3.2 Expérience numérique

Nous proposons d'étudier la dynamique des rendements logarithmiques en simulant les changements d'opportunité de profits parmi les agents financiers. En considérant les hypothèses d'Osborne, si l'opportunité moyenne de profits à long terme du système converge vers 0,5 (c'est-à-dire la probabilité marginale de profits pour un pari), alors la moyenne à long terme des rendements logarithmiques doit converger vers 0, avec une symétrie significative au sein de sa distribution en raison de l'homogénéité des opportunités entre agents. Nous testons ces hypothèses.

3.3.2.1 Système conservatif

Nous considérons le système constituant un ensemble d'agents financiers comme un macro-état Ω où chaque agent décide de parier à la hausse comme à la baisse sur les marchés financiers en fonction d'un ensemble d'informations données. Nous décrivons chaque agent comme un micro-état \mathcal{M}_i de Ω avec une probabilité de profits donnée $p_{\mathcal{M}_i}$, où $i = 1, \dots, n$, avec n le nombre de micro-états. Ce nombre est fixe et restera le même jusqu'à la fin de l'expérimentation (système conservatif). Initialement, nous supposons, comme Osborne, que les agents sont homogènes et partagent la même opportunité de profits. Cependant, au fur et à mesure que les paris avancent, chaque

agent met à jour sa probabilité de profits en fonction de sa série de résultats précédente. Ainsi, maximiser leur opportunité de profits repose sur la maximisation de leur probabilité postérieure d'observer un profit sur leur prochain pari, compte tenu de leurs résultats historiques :

$$P(\text{win}|\mathcal{M}_i) = \frac{P(\text{win}) * P(\mathcal{M}_i|\text{win})}{P(\text{win}) * P(\mathcal{M}_i|\text{win}) + P(\text{loss}) * P(\mathcal{M}_i|\text{loss})} \quad (3.57),$$

$P(\text{win}) = P(\text{loss})$ est la probabilité marginale de profit et de perte respective.

La figure A.1 (p.158) fournit une illustration simplifiée de la simulation numérique. Nous sommes en présence de six macro-états dans lesquels, à chaque étape au moins, une transaction entre deux micro-états se produit. Chaque macro-état est constitué de cinq micro-états indépendants {A, B, C, D, E}, et nous supposons à l'étape 1 être sous la condition d'Osborne, c'est-à-dire que nous avons initialement une homogénéité parfaite d'opportunité de profits entre les micro-états. La figure A.1 (p.158) doit être lue en suivant les flèches, de sorte que la transaction se produisant à l'étape 1 soit suivie de la transaction se produisant à l'étape 2, et ainsi de suite...

Nous représentons la probabilité postérieure de chaque micro-état en utilisant une couleur donnée. Si la couleur est la même pour différents micro-états, cela signifie qu'ils sont homogènes. La couleur des micro-états change à chaque étape, car chaque micro-état met à jour sa probabilité postérieure en fonction du résultat du dernier pari. Le tableau A.1 (p.159) met l'accent sur cette évolution de la probabilité postérieure pour chaque micro-état, au fur et à mesure que les paris progressent dans le temps.

La simulation numérique est basée sur la même logique. Nous commençons avec 750 micro-états qui ont initialement la même probabilité de profits. A chaque itération, au moins un pari entre deux micro-états se produit conduisant à une mise à jour de leur probabilité de profits postérieurs respective. Nous exécutons le modèle pour 10 000 itérations et déterminons, pour chacune d'entre elles, la moyenne lissée de la probabilité de profits postérieurs du macro-état, la distribution de la probabilité postérieure du macro-état, l'évolution de l'entropie du macro-état, et enfin l'évolution

des probabilités de profits postérieurs de chaque micro-état. Le macro-état, qui définit l'équilibre du système, est celui constitué du nombre maximal de micro-états identiques, à savoir celui qui maximise son entropie. En d'autres termes, plus le système est constitué de micro-états avec une probabilité profits égale, plus le système tend à converger vers son équilibre et à maximiser son entropie.

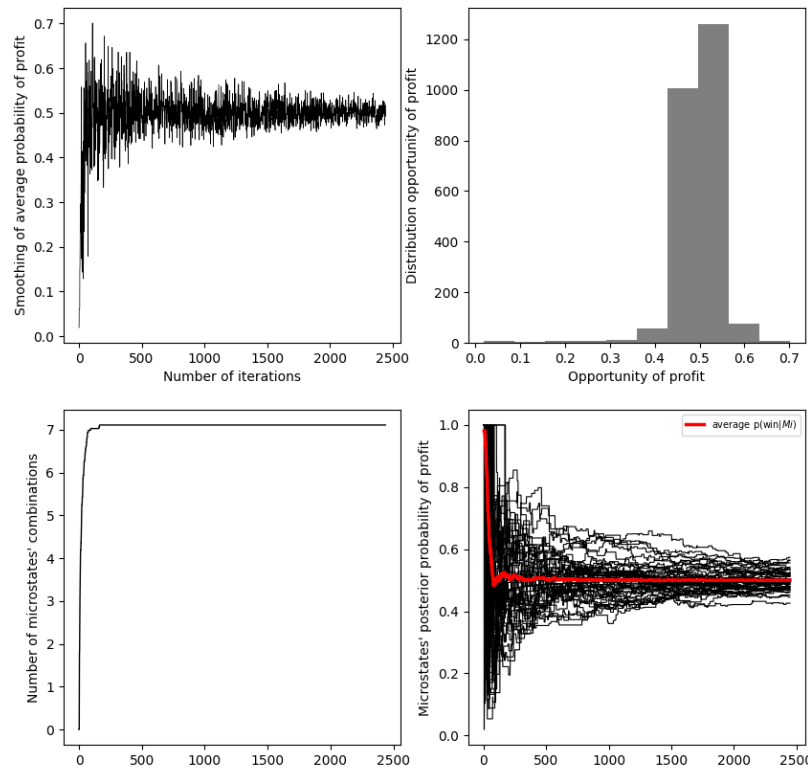


Figure 3. 5 Expérience sous l'hypothèse d'un système conservatif. Graphique supérieur gauche : moyenne lissée de la probabilité de profit postérieur du macro-état. Graphique supérieur droite : distribution de la probabilité de profit du macro-état. Graphique inférieur gauche : évolution du nombre de combinaisons possibles de micro-états avec une probabilité marginale égale à 0,5. Graphique inférieur droite : évolution de la probabilité postérieure d'opportunité de profit de chaque micro-états. La courbe rouge représente la probabilité moyenne de l'ensemble des micro-états.

La figure 3.5 met l'accent sur les résultats statistiques de la première expérience. La distribution des opportunités de profits est fortement concentrée autour de sa moyenne, à peu près égale à 0,5, ce qui signifie qu'au fil du temps, la probabilité moyenne de profits du système converge vers une probabilité marginale de profits d'un jeu non biaisé. Nous faisons le même constat sur l'évolution de l'opportunité de profits des micro-États. Au fil du temps, ils convergent tous vers 0,5. La probabilité marginale

de profits est atteinte lorsque le nombre de micro-états avec équiprobabilité de profits a atteint son maximum, ce qui signifie que l'entropie du système est maximisée. En raison de l'hétérogénéité des informations disponibles entre les micro-états, nous constatons la présence d'asymétries au sein de la distribution des opportunités de profits. Dans le tableau 3.1, nous mesurons une asymétrie de -1,45, ce qui contredit la théorie d'Osborne selon laquelle la symétrie et l'équilibre du système sont atteints lorsque la distribution des opportunités de profits converge vers une distribution normale. Cependant, un kurtosis de 3,21 indique que la distribution n'est pas loin d'être platikurtique, signifiant qu'il n'y a que très peu de valeurs aberrantes.

Données simulées	Moyenne	Mediane	Skewness	Kurtosis
Opportunité de profit	0.468	0.487	-1.45	3.21

Tableau 3. 1 Résultats statistiques des données simulées et de leur distribution (Cas d'un système conservatif).

Nous comparons la distribution des opportunités de profits issus de la simulation à d'autres distributions, à savoir la distribution normale, la distribution t-student, la distribution de Cauchy, la distribution de Pearson et la distribution de Laplace. Nous utilisons ces distributions car, comme Mandelbrot et Fama [78, 79], nous supposons que les distributions qui représentent les variations de prix sont intermédiaires entre une distribution de Cauchy et une distribution normale. La figure 3.6 et le tableau 3.2 montrent nos résultats. Nous comparons toutes les distributions en utilisant la somme des erreurs quadratiques et montrons que la distribution t-student doit être préférée pour expliquer la dynamique qui émerge de la distribution des opportunités de profits entre agents. Nous calculons une p-value avec un niveau de significativité à 5 % par la méthode « bootstrap » en simulant plusieurs échantillons. Les résultats sont unanimes : parmi toutes les distributions testées, la distribution t-student reste la plus significative avec une p-value convergeant vers 0,429. En revanche, nous montrons que nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle la distribution des opportunités de profits est normalement distribuée. Les résultats sont présentés dans le tableau 3.2 et la figure A.3 (p.156).

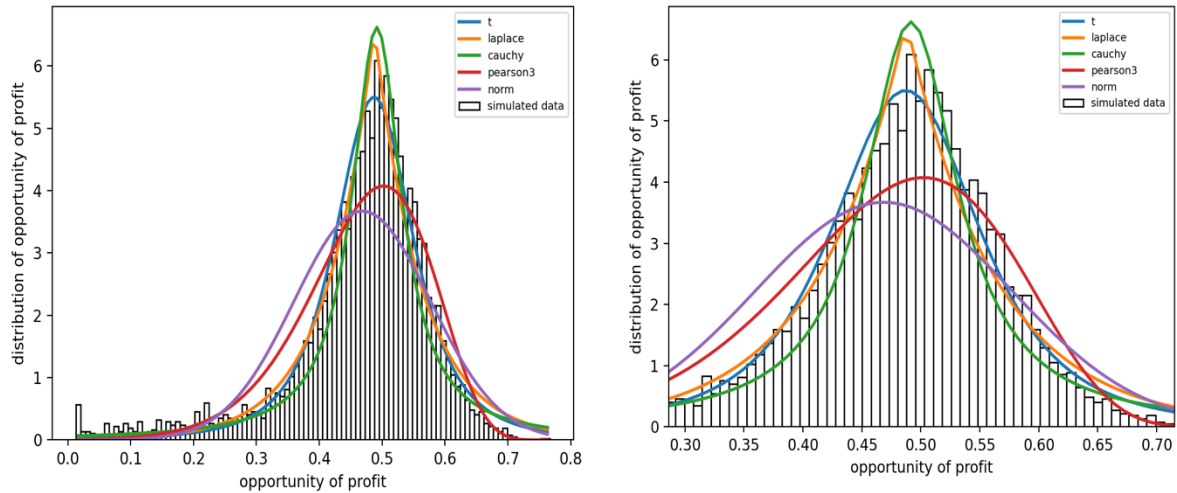


Figure 3. 6 Cas d'un système conservatif. A gauche : répartition de l'opportunité de profits sous forme d'histogrammes. Modélisation de l'opportunité de profit selon différentes statistiques (bleu : t-étudiants, jaune : Laplace, vert : Cauchy, rouge : Pearson et violet : norme). A droite : même figure mais agrandie.

Distribution model	SSE	aic	P-value	Paramètres estimés
t-student	7.80	172.95	0.429	$v: 2.72, \mu: 0.49, \sigma: 0.66$
Laplace	12.80	159.65	0.284	$\mu: 0.49, \sigma: 0.75$
Cauchy	19.42	141.01	0.29	$\mu: 0.49, \sigma: 0.048$
Pearson	27.92	245.10	0.387	$v: -0.66, \mu: 0.47, \sigma: 0.10$
Normal	57.02	197.07	0.127	$\mu: 0.47, \sigma: 0.11$

Tableau 3. 2 Comparaison de différentes modèles statistiques avec la distribution des opportunités de profits simulés. Méthode de la somme des erreurs au carré es utilisée. Le modèle de distribution qui modélise les données observées avec le SSE le plus faible doit être préféré.

3.3.2.2 Système dissipatif

Nous avons précédemment considéré le système comme fermé, constitué d'un nombre fixe d'agents financiers. Nous considérons maintenant le système comme un système de type « coarse-grained », constitué d'un ensemble de groupes d'agents financiers homogènes interagissant avec d'autres groupes hétérogènes. Le système est également dissipatif, permettant à de nouveaux agents financiers d'entrer ou de sortir du système. Nous considérons également les entrées et sorties d'agents comme indépendantes du temps, ce qui conduit à des dynamiques spatio-temporelles inhomogènes. La figure A.2 (p.160) propose une illustration sur laquelle est construit

le modèle de simulation. On considère maintenant le macro-état Ω constituant un ensemble inhomogène de macro-états locaux $(\omega_1, \omega_2, \omega_3 \dots \omega_N)$ plus ou moins hétérogènes, interagissant entre eux, constituant des micro-états respectifs de tailles différentes tels que $\mathcal{M}_1 \subset \omega_1$ et $\mathcal{M}_1 = \{\mathcal{M}_{1i}\}$ avec $i = 1, \dots, n_1$; n_1 correspond au nombre de micro-états \mathcal{M}_1 . Nous considérons cette représentation comme plus réaliste que le système conservatif. Dans cette expérimentation, les agents financiers mettent à jour leur opportunité de profit selon la même inférence que celle décrite dans l'expérience précédente.

La simulation numérique modélise la dynamique de quatre macro-états locaux ω_N ($N=4$) constituant respectivement 750, 225, 150 et 425 micro-états. Le nombre de micro-états est un choix arbitraire et n'a aucun impact sur la dynamique globale, seulement sur sa vitesse de convergence vers l'équilibre. Ces micro-états n'interagissent pas simultanément avec le système, mais interagissent avec des intervalles de temps différents. Par conséquent, chaque micro-état met à jour sa probabilité de profit à une vitesse différente. Comme dans l'expérience précédente, nous exécutons le modèle pour 10 000 itérations et illustrons les résultats de la simulation dans la figure 3.7. Nous montrons que la dynamique du système peut être vue comme une superposition de différentes statistiques locales correspondant au comportement respectif de l'état de chaque macro-état local. Nous observons que chaque macro-état local converge vers la même probabilité marginale de profit de 0,5, mais selon une vitesse différente. Cette vitesse de convergence dépend du nombre d'interactions avec le système et, par conséquent, du nombre de micro-états qui le constituent. Plus le nombre de micro-états est grand, plus la convergence est lente.

De plus, cette variation de vitesse dans la dynamique conduit à plus d'asymétries dans la distribution des opportunités de profit du système car elle exacerbe son caractère hétérogène. Le tableau 3.3 met en avant les statistiques concernant les données simulées. Le kurtosis est égal à 9,11 et le skewness est égal à -2,33, les deux étant bien supérieurs à ceux de l'expérience précédente. Cela correspond à une distribution leptokurtique avec plus de données loin de l'équilibre du système. Par

conséquent, les queues de distribution s'approchent de zéro plus lentement que sous une distribution gaussienne, produisant plus de valeurs aberrantes que la distribution normale et contredisent la théorie d'Osborne. Nous ajoutons que cette simulation est une simplification des hétérogénéités et des degrés de liberté du marché réel, ce qui signifie que si nous pouvions simuler les changements réels d'opportunité de profit des agents financiers, nous pouvons supposer que nous observerions une distribution beaucoup plus éloignée de l'équilibre et avec plus d'asymétries.

Données simulées	Moyenne	Mediane	Skewness	Kurtosis
Opportunité de profit	0.485	0.496	-2.33	9.11

Tableau 3. 3 Résultats statistiques des données simulées et de leur distribution (Système dissipatif).

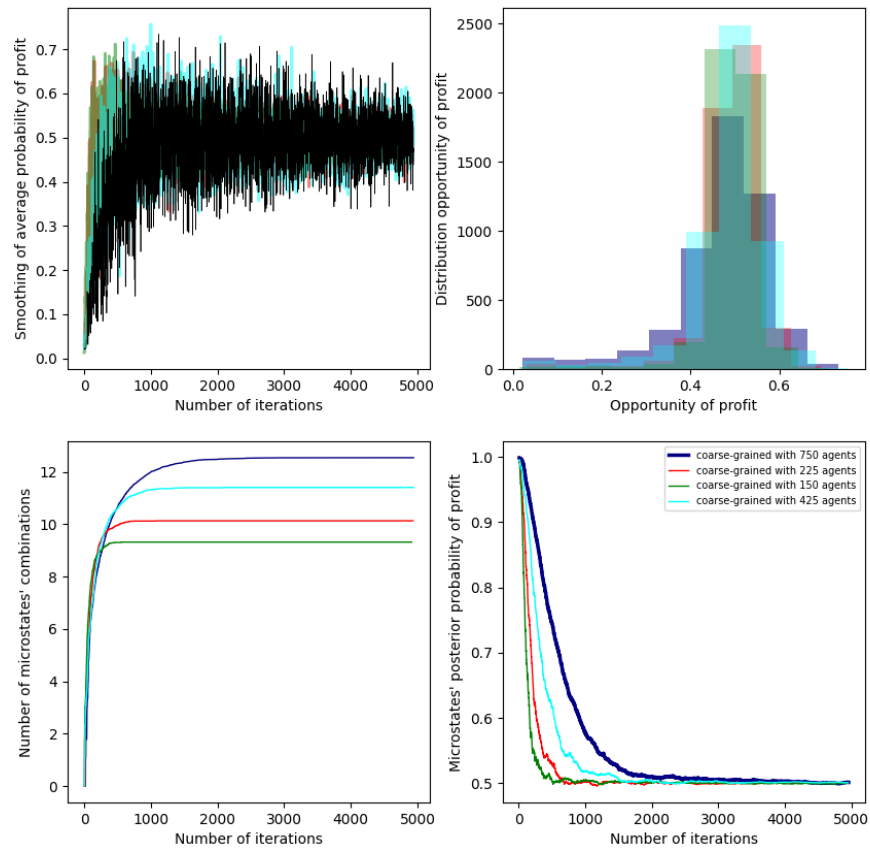


Figure 3. 7 Expérience sous l'hypothèse d'un système dissipatif. Graphique supérieur gauche : moyenne lissée des probabilités postérieures de profit des macro-états locaux (dit « coarse-grained »). Graphique supérieur droite : distribution de la probabilité postérieure des opportunités de profits de chaque macro-état local. Graphique inférieur gauche : évolution du nombre de micro-états avec une probabilité marginale de 0.5 pour chaque macro-état local. Graphique inférieur droite : évolution de la probabilité postérieure de profit moyenne des macro-états locaux.

La figure 3.8 et le tableau 3.4 mettent en avant les résultats de la comparaison entre les données simulées d'opportunité de profit du macro-état global et la distribution normale, la distribution t-student, la distribution de Cauchy, la distribution de Pearson et la distribution de Laplace. Comme dans l'expérience précédente, nous avons comparé toutes les distributions en utilisant la somme des erreurs quadratiques et montrons que la distribution t-student devrait être préférée pour expliquer la dynamique d'opportunité de profit entre les agents, c'est-à-dire infime la variation des rendements logarithmiques. Nous calculons une p-value avec un niveau de significativité à 5 % par la méthode « bootstrap » en simulant plusieurs échantillons. Les résultats sont unanimes : parmi toutes les distributions testées, la distribution t-student reste la plus significative avec une p-value convergeant vers 0,466. Nous rejetons également l'hypothèse nulle selon laquelle la distribution des opportunités de profit est normalement distribuée. Les résultats sont présentés dans le tableau 5 et la figure A.3 (p.160).

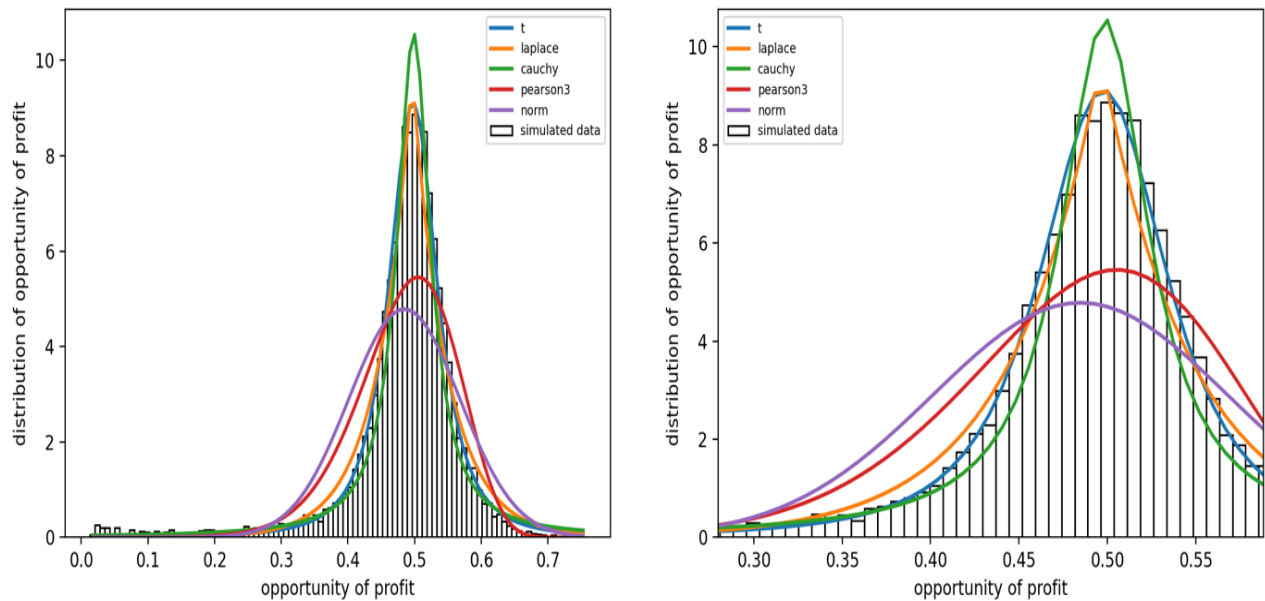


Figure 3. 8 Superposition de la distribution des macro-états local d'un système dissipatif. A gauche : répartition de l'opportunité de profit du système sous forme d'histogrammes. Modélisation de l'opportunité de profit selon différentes statistiques (bleu : t-étudiants, jaune : Laplace, vert : Cauchy, rouge : Pearson et violet : norme). A droite : même figure mais agrandie.

Distribution model	SSE	aic	P-value	Parameters fitting
t-student	2.77	283.36	0.466	$v: 2.12, \mu: 0.50, \sigma: 0.04$
Laplace	13.42	339.35	0.08	$\mu: 0.49, \sigma: 0.05$
Cauchy	20.69	206.17	0.348	$\mu: 0.50, \sigma: 0.03$
Pearson	91.28	575.71	0.398	$v: -0.54, \mu: 0.48, \sigma: 0.076$
Normal	142.78	496.66	0.018	$\mu: 0.48, \sigma: 0.083$

Tableau 3. 4 Comparaison de différents modèles statistiques avec la distribution des opportunités de profits simulés. La méthode de la somme des erreurs au carré est utilisée. Le modèle de distribution qui modélise les données observées avec le SSE le plus faible doit être préféré.

Nous montrons que, dès que le système devient moins dissipatif ¹²avec des micro-états en interaction qui deviennent moins hétérogènes, chaque macro-état local converge vers la probabilité d'équilibre de profits de 0,5. Par conséquent, la vitesse de convergence du système dépend de son degré de dissipation. On peut supposer que, sur les marchés financiers, cette dissipation correspond au niveau de liquidité du marché, au volume de transactions et, à fortiori, au nombre d'agents financiers hétérogènes interagissant sur le marché. Nous montrons également que la probabilité postérieure d'opportunité de profits de chaque macro-état local n'évolue pas selon la même dynamique. Certains convergent également plus rapidement vers leur état d'équilibre. Ceci est dû au fait que, dans notre modèle, chaque macro-état local et à fortiori micro-état ne partagent pas la même information et, par conséquent, n'actualisent pas leur probabilité de profits selon la même dynamique. Une conséquence est que chaque macro-état local est défini par une distribution respective d'opportunités de profits avec une variance respective, comme le montrent la figure 3.9 et le tableau 3.5. En revanche, en ce qui concerne la distribution de la probabilité postérieure d'opportunité de profits de chaque macro-état local, nous montrons par p-value que la distribution t-Student est également celle qui est la plus significative.

Les résultats sont présentés dans la figure A.4 (p.161).

¹² Parce que nous ne pouvons pas exécuter le modèle avec un nombre infini de nouveaux macro-états locaux, au fur et à mesure des itérations, le système devient plus conservateur car il y a moins d'interactions avec de nouveaux micro-états.

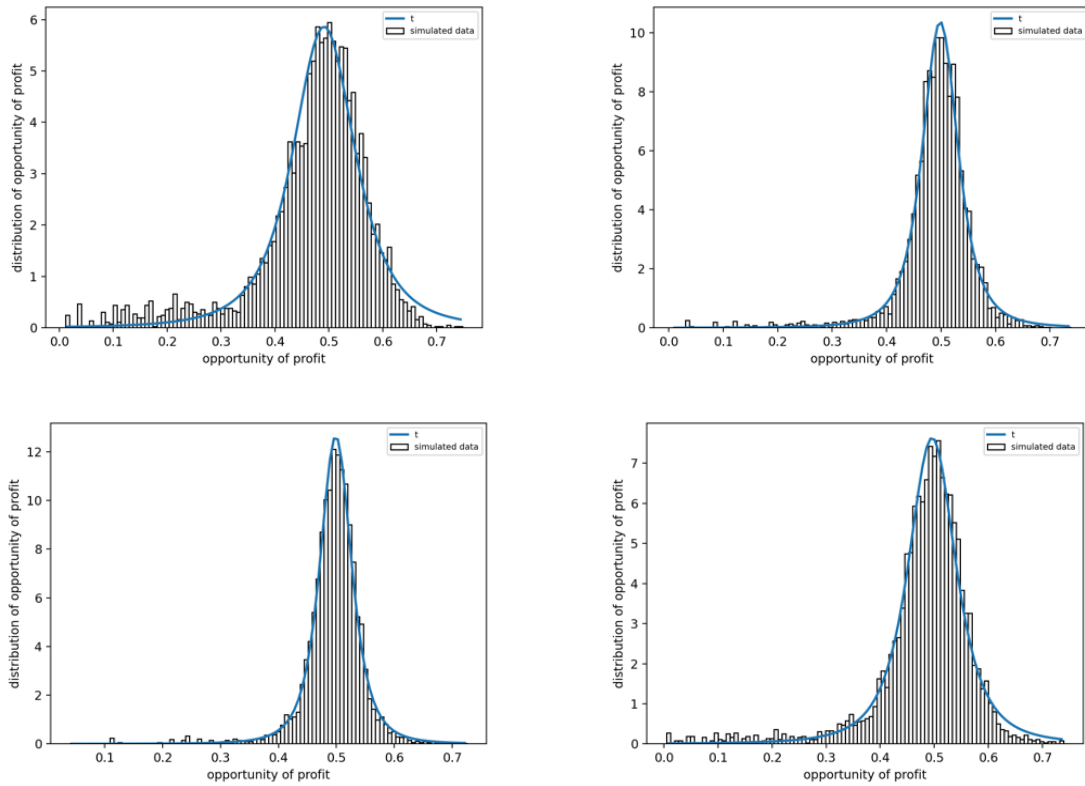


Figure 3. 9 Distribution de probabilité des opportunités de profits de chaque macro-état local en histogramme avec la distribution de t-student en bleu. En haut à gauche : macro-état local ω_1 avec 750 micro-états. En haut à droite : macro-état local ω_2 avec 225 micro-états. En bas à gauche : macro-état local ω_3 avec 150 micro-états. En bas à droite : macro-état local ω_4 avec 425 micro-états.

Distribution model	Mean, variance	skew	kurtosis	t-student fitting	SSE
$\omega_1 = 750 \text{ agents}$	$\mu: 0.47, \sigma^2 : 0.011$	-1.5	3.25	$v: 2.48, \mu: 0.49, \sigma: 0.06$	10.45
$\omega_2 = 225 \text{ agents}$	$\mu: 0.49, \sigma^2 : 0.005$	-2.43	11.98	$v: 2.51, \mu: 0.49, \sigma: 0.03$	11.06
$\omega_3 = 150 \text{ agents}$	$\mu: 0.47, \sigma^2 : 0.011$	-2.31	13.35	$v: 2.75, \mu: 0.49, \sigma: 0.029$	6.88
$\omega_4 = 425 \text{ agents}$	$\mu: 0.48, \sigma^2 : 0.008$	-2.01	6.82	$v: 2.38, \mu: 0.49, \sigma: 0.047$	7.35

Tableau 3. 5 Résultats Statistiques de l'estimation des paramètres de la distribution des macro-états locaux avec la loi t-student.

Pris individuellement, chaque macro-état local peut ne pas décrire une distribution leptokurtique de la même manière que la distribution issue de l'expérience d'un système conservatif. Globalement, la distribution du système consiste en une superposition des statistiques de chaque macro-état local caractérisé par une variance différente et une moyenne connue. Cette moyenne finit par converger vers une probabilité marginale de profits de 0,5. Nous affirmons que cette variance qui est non constante entre différents macro-états locaux, est source à l'échelle globale d'émergence

de plus d'asymétries et d'excès de kurtosis au sein de la distribution observée (Figure 3.8).

Finalement, dans un système dissipatif, il existe suffisamment de preuves pour rejeter la distribution normale proposée par Osborne se caractérisant notamment par une parfaite symétrie de profits entre agents. En raison de la superposition de statistiques, nous montrons que l'autre hypothèse majeure dans la dérivation du modèle d'Osborne concernant une variance constante des rendements logarithmiques n'est pas valide. Nous montrons dans notre deuxième expérience que la variance σ^2 représente le degré d'activité des interactions des agents financiers et varie selon l'espace et le temps. Cette variance fluctuante à l'échelle globale due à la superposition de statistiques est ce que nous traiterons dans les sections suivantes. En effet, nous postulons que la complexité d'une série financière mise en lumière dans le Chapitre 2 ainsi que sa dépendance spatio-temporelle sont la conséquence de ce processus de superposition de statistiques qui se traduit par une variance non constante. Ainsi, nous considérerons dorénavant que la variance des rendements logarithmiques d'un cours boursier est une variable aléatoire de distribution $g(\sigma^2)$.

3.3.3 Volatilité stochastique et super-statistique

Comme indiqué dans la section précédente, nous considérons maintenant la variance des changements de prix d'un cours boursier comme une variable aléatoire avec une fonction de distribution $g(\sigma^2)$. Par conséquent, la distribution donnée par Osborne dans l'équation (3.56) doit être réinterprétée comme une distribution de probabilité conditionnée à la variance σ^2 , telle que l'équation (3.56) devient :

$$P(x|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \quad (3.58)$$

Et $P(x)$ devient la distribution de probabilité de x , qui prend en compte le caractère aléatoire de σ^2 , de sorte que l'on obtient $P(x)$ en résolvant l'équation suivante :

$$P(x) = \int_0^\infty P(x|\sigma^2) * g(\sigma^2) d\sigma^2 \quad (3.59),$$

avec $0 \leq \sigma < \infty$. Dans l'équation (3.59), $g(\sigma^2)$ représente le préjugé de la distribution du paramètre inconnu σ^2 . Comme déjà discuté par Raiffa et Schlaifer [102], une approche efficace pour déterminer la dynamique de σ^2 est une alternative bayésienne où la vraisemblance de $P(x|\sigma^2)$ est définie par une distribution normale standard de moyenne μ connu et de variance inconnue σ^2 où le conjugué de $g(\sigma^2)$ pour un échantillon supposé normalement distribué est théoriquement accepté comme étant une distribution inverse gamma définie par l'équation (B.6, p.163).

Nous supposons également qu'il n'est pas nécessaire d'avoir un préjugé sur la moyenne μ puisqu'à partir de nos expériences précédentes, nous montrons sa convergence vers l'équilibre conduisant à une moyenne à long terme des rendements logarithmiques convergeant vers zéro. D'autres études¹³ montrent des preuves empiriques que la variabilité de la moyenne des rendements logarithmiques n'est pas aussi significative que la fluctuation de la variance. Une étude plus récente de Tao Ma et R.A. Serota [76] alternative à l'inférence Bayésienne conforte les résultats de Raiffa et Schlaifer concernant le choix de la distribution inverse gamma pour modéliser la distribution de la variance σ^2 . En revanche ils montrent également que d'autres distributions peuvent être considérées comme candidates, à savoir la distribution log-normale (LN), gamma (Ga) et la distribution généralisée inverse gamma (GIGa), car toutes appartiennent à la même famille de distributions. Ils concluent également de manière similaire à nos résultats expérimentaux précédents, à savoir que la distribution t-student est la meilleure candidate pour modéliser les rendements logarithmiques des cours de l'indice boursier Standard&Poor500 et soutiennent que les rendements peuvent être modélisés de manière significative par le produit de la distribution de la variance $g(\sigma^2)$, avec une distribution normale de moyenne nulle et de variance unitaire.

Leurs conclusions et les nôtres, tirées de nos deux expérimentations précédentes, se confortent et renforcent l'alternative de l'inférence bayésienne pour modéliser la

¹³ Dyer, James C, Random walk in Australian share prices: a question of efficient capital markets, Australian economic papers, 1976-12, Vol.15 (27), p.186-200.

distribution de la variance. Cependant, dans leur étude, Tao Ma et R.A. Serota [76] ne s'intéressent pas à la dépendance de la distribution de la variance à l'échelle temporelle. Si, dans nos expériences ci-dessus, la fonction de distribution $g(\sigma^2)$ représente la variation future d'opportunité de profits des agents ou de groupe d'agents financiers (c'est-à-dire, in fine la variation des rendements espérés), on peut supposer que la nature de cette fonction peut changer en fonction du temps puisque l'espérance des opportunités de profits sont également variables.

Dans un premier temps, nous proposons dans la suite de ce Chapitre de modéliser la complexité du système émergent du paramètre fluctuant qu'est la variance σ^2 sur différentes échelles de temps intra-journalières.

Pour la modélisation de la variance, nous proposons, par inférence Bayésienne, de comparer deux modèles standards en super statistiques, à savoir le modèle inverse-gamma et le modèle log-normal. Puis une fois les modèles calibrés, dans une démarche similaire de celle de Tao Ma et R.A Serota, nous prouvons par la « p-value » qu'en effet la distribution des rendements logarithmiques observables peut être simulée de manière significative par le produit de la distribution normale et de la super-statistique inverse gamma sur l'échelle temporelle « minute ». Sur de plus grandes échelles intra-journalières, le produit de la distribution normale et log-normal devient plus significatif. En revanche, nous montrons également qu'à mesure que l'échelle temporelle augmente, ces deux modèles perdent en significativité dans la description de la distribution des rendements logarithmiques.

Ainsi, dans une dernière partie nous montrons l'évidence d'une transition de statistique de la complexité des séries financières ; transition de haute fréquence vers basse fréquence. Pour ce faire nous ferons nos analyses statistiques sur des échelles temporelles plus larges, notamment « hebdomadaire » et « mensuelle » et selon d'autres super-statistiques de la même famille, notamment la distribution gamma.

3.4 Étude intra-journalière de la complexité des séries financières : comparaison de super-statistiques par inférence Bayésienne

Comme explicité ci-dessus et mis en avant dans l'annexe B (p.162-163), à partir d'une inférence bayésienne et en partant de l'hypothèse que la vraisemblance marginale du modèle sachant les données \mathcal{D}_n est une distribution de probabilité gaussienne avec une moyenne nulle μ et avec une variance σ^2 considéré étant un paramètre libre, nous montrons que la paramétrisation la plus pratique qui maximise la probabilité postérieure est de choisir une probabilité conjuguée $g(\sigma^2) = \pi_{\alpha,\beta}(\sigma^2)$ qui est régie par une distribution inverse gamma de paramètres α et β .

Cependant, même si d'après une approche bayésienne, la distribution de densité de probabilité inverse gamma semble être celle qui maximise notre credence concernant la probabilité postérieure du paramètre fluctuant σ^2 , il est également possible qu'en ce qui concerne des échelles de temps spécifiques, des distributions de même famille liées à la fonction inverse gamma pourraient être tout aussi pertinentes en termes de maximisation de la probabilité postérieure. Par conséquent, nous proposons dans cette section de comparer deux modèles super statistiques standards sur différentes échelles de temps intra-journalières. Le test de facteur de Bayes est utilisé pour définir quel modèle est le plus pertinent et pour quelles échelles de temps.

3.4.1 Analyse de données

Nous utilisons comme données empiriques l'indice « indice Américain Standard&Poor500 » en s'appuyant sur les travaux de Lanford en 1973 [72]. Il soutient que le niveau de pertinence de l'utilisation de la mécanique statistique pour expliquer les comportements d'un système dépend de son degré de liberté. Son approche repose sur le fait que la probabilité d'erreurs dans la détermination statistique des comportements complexes d'un système à un niveau macroscopique est inversement proportionnelle au nombre de degrés de liberté qui caractérise le système. Plus on a de particules microscopiques ou de dynamiques d'équilibre local avec des caractéristiques statistiques respectives, plus le système de dynamique de non-équilibre à l'échelle

macroscopique tend à converger vers un comportement dynamique relativement proche d'une dynamique moyenne de chaque dynamique d'équilibre local. Par conséquent, nous considérons l'indice S&P500 comme un système complexe constitué d'un grand nombre d'agents hétérogènes dans le temps et dans l'espace avec différentes opportunités de profits.

Pour cette étude, nous utilisons comme données les rendements logarithmiques de l'indice sur l'année complète 2020, soit sur 52 semaines. Ce choix est fortement motivé par la volatilité exacerbée des marchés qui s'était produite pendant la période de pandémie de Covid-19. Nous étudions la dynamique des séries temporelles selon quatre différentes échelles de temps intra-journalières : « minute », « heures », « 4 heures » et « journalière ». On définit le logarithme des prix comme étant le système étudié, l'évolution des prix logarithmes comme étant la dynamique du système et la fluctuation du système comme étant la variance σ^2 et $g(\sigma^2) = \pi_{\alpha,\beta}(\sigma^2)$ définissant la distribution de probabilité de cette variance sur chaque échelle de temps étudiée. De plus, comme mentionné dans la section 2, nous considérons que la vraisemblance des données observées \mathcal{D}_n du système (c.-à-d., les rendements logarithmiques) convergent vers une distribution normale. L'équation de la vraisemblance marginale est définie par l'équation 3.59. Nous définissons le processus de volatilité comme étant la variation absolue des rendements logarithmiques telle que : $|r_t| = |\log(P_{t+1}) - \log(t)|$ où $t = t_0, t_1, \dots, t_n$ avec la taille de n en fonction de l'échelle de temps. Figure 3.10, on propose une représentation sur les quatre échelles de temps d'études. Nous constatons au premier abord la présence d'hétéroscédasticité, suggérant des variations significatives de la variance. De même, le phénomène de concentration de la volatilité a tendance à diminuer à mesure que les échelles de temps augmentent.

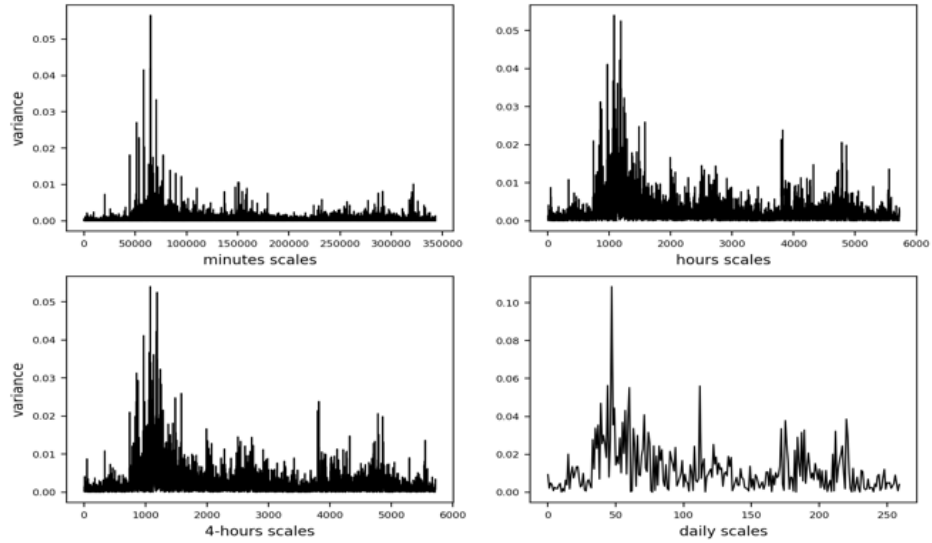


Figure 3. 10 Variation absolue des rendements logarithmiques du S&P500 selon différente échelle temporelle. En haut à gauche : « minute ». En haut à droite : « heure ». En bas à gauche : « 4-heures ». En bas à droite : « journalière ».

Nous quantifions statistiquement le constat précédent en déterminant l'exposant de queue de la loi de puissance des rendements logarithmiques absolus de l'indice Standard&Poor500. L'exposant d'une queue de loi de puissance peut être calculé par l'équation suivante :

$$\mathbf{CDF}(\omega) = \mathbf{1} - \mathbf{CDF}(\omega) = \int_{\omega}^{+\infty} \mathbf{PDF}(\omega) d\omega \quad (3.60)$$

$$\text{si } \mathbf{PDF}(\omega) \propto \omega^{-1-k} \text{ with } \omega \gg 1,$$

$$\text{alors : } \log(\mathbf{CCDF}(\omega)) \propto \text{constant} - k \log(\omega) \quad (3.61),$$

avec CDF la fonction de répartition, PDF la probabilité de distribution, et CCDF la fonction de distribution cumulative complémentaire.

Ainsi, si le logarithme de CCDF est une fonction linéaire de $\log(\omega)$, avec ω correspondant aux différentes valeurs de la variance, alors on en déduit que la distribution de ω se caractérise par une loi de puissance.

La figure 3.11 donne une représentation graphique « log-log » des queues de données empiriques pour des échelles de temps données. Il est clair que les régressions linéaires gagnent en significativité dès lors que les échelles de temps diminuent (en particulier sur l'échelle « minute ») et deviennent progressivement moins significatives à mesure que les échelles de temps augmentent. On en déduit que la distribution de la variance a tendance à se caractériser par une loi de distribution à queue lourde et qu'elle est corrélée à une distribution de loi de puissance à mesure que les échelles de temps diminuent.

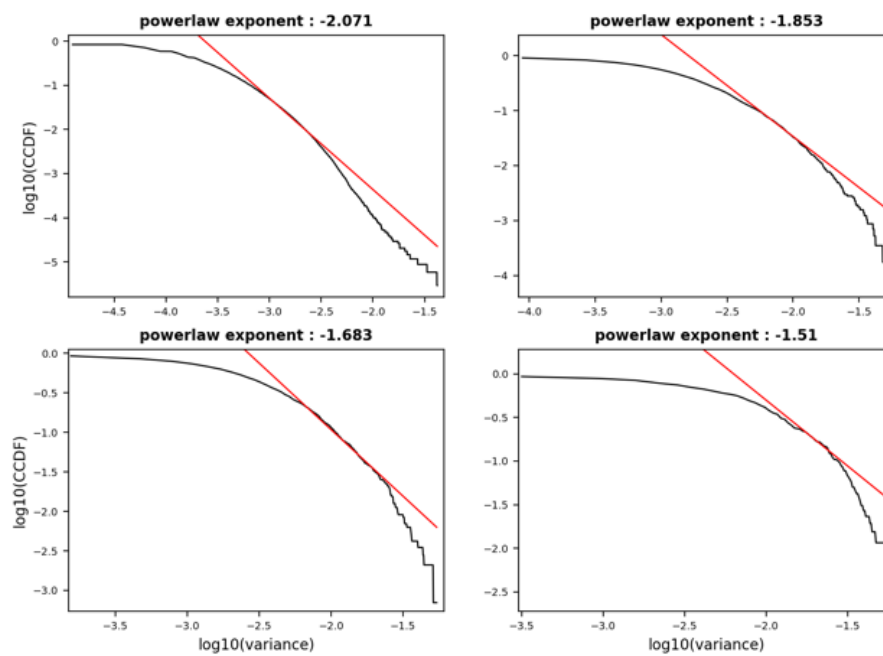


Figure 3. 11 Graphique log-log de CCDF en fonction de la variance. Courbes noires : tracé log-log. Droite rouge : Régression linéaire du $\log_{10}(\text{CCDF})$. En haut à gauche : période = « minute », $k = -2.071$. En haut à droite : période = « heure », $k = -1,853$. En bas à gauche : période = « 4 heures », $k = -1,683$. En bas à droite : période = « journalière », $k = -1.051$.

Sur les échelles de temps « minute », « heure » et « 4 heures », le test de stationnarité Kwiatkowski-Phillips-Schmidt-Shin-KPSS (tableau 3.6) montre la présence d'un processus non stationnaire (présence d'une racine-unitaire) indiquant une certaine indépendance de la variance sur le temps en tant que variable fluctuante. Nous constatons également que le degré de stationnarité augmente à mesure que l'échelle de temps augmente. Finalement, sur l'échelle « journalière », nous ne pouvons pas rejeter

l'hypothèse nulle (c'est-à-dire l'hypothèse de stationnarité) indiquant une fluctuation de la variance dépendante du temps.

	Minutes	Hours	4-Hours	Daily
Test Statistic	45.72	2.55	1.013	0.3632
P-value	0.01	0.01	0.01	0.093
Valeur critique (10%)	0.347			
Valeur critique (5%)	0.463			
Valeur critique (2.5%)	0.5754			
Valeur critique (1%)	0.739			

Tableau 3. 6 Test de stationnarité Kwiatkowski-Phillips-Schmidt-Shin (KPSS) sur les rendements logarithmiques absolus. L'hypothèse nulle : la série temporelle est stationnaire. Hypothèse alternative : la série temporelle a une racine unitaire (non stationnaire). Si la statistique de test est supérieure à la valeur critique, nous rejetons l'hypothèse nulle.

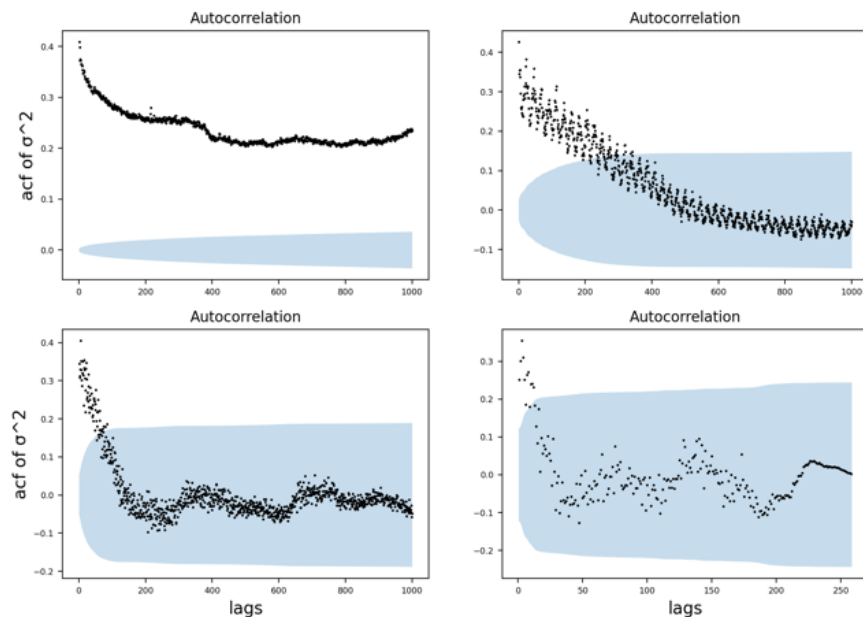


Figure 3. 12 Fonction d'autocorrélation des valeurs de la variance du Standard&Poor500. En haut à gauche : « minutes », En haut à droite : « heures », En bas à gauche : « 4 heures », En bas à droite : « journalière ». La courbe bleue est la zone d'intervalle de confiance de 5%.

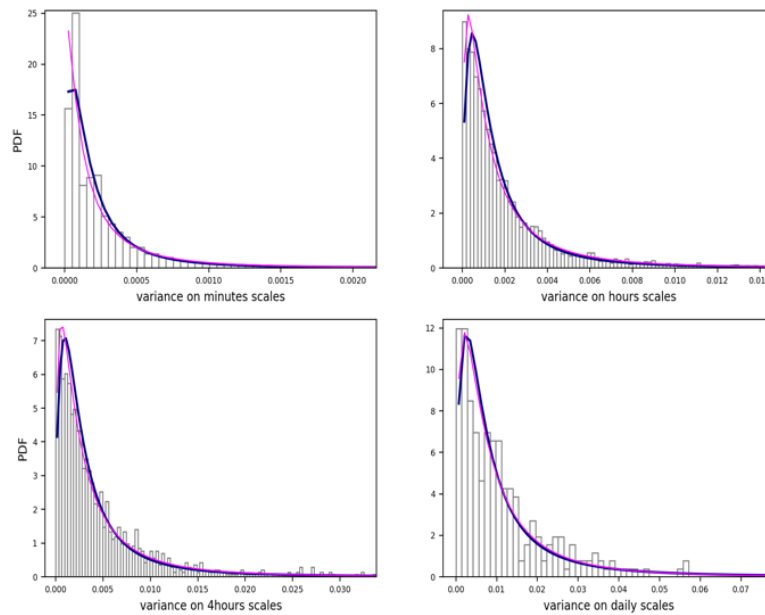


Figure 3. 13 Histogramme de la variance du Standard&Poor500. En haut à gauche : séries chronologiques sur l'échelle « minute ». En haut à droite : séries chronologiques sur l'échelle « heure ». En bas à gauche : séries chronologiques sur l'échelle « 4 heures ». En bas à droite : séries chronologiques sur l'échelle « journalière ». Ligne bleue : distribution des données observées avec la distribution inverse gamma. Ligne rose : distribution des données observées avec la distribution log-normal.

La figure 3.12 met l'accent sur le phénomène de mémoire longue et de dépendance long-terme selon chaque échelle de temps. La fonction d'autocorrélation avec un intervalle de confiance de 5 % montre que, sur des échelles de temps très courtes (c'est-à-dire « minute »), la fonction d'autocorrélation de la variance σ^2 décroît plus lentement qu'une décroissance exponentielle. Cette dernière va de pair avec la valeur de l'exposant de la loi de puissance déterminée ci-dessus et mise en avant par la figure 2. À mesure que les échelles de temps augmentent, la fonction d'autocorrélation tend vers une décroissance de type exponentielle, de sorte que, sur l'échelle « journalière » ou même « 4 heures », le phénomène de mémoire s'estompe. La figure 3.13 montre la distribution en histogramme de la variance du Standard&Poor500 sur les échelles de temps d'études. Sur des hautes échelles à plus hautes fréquences telles que « minute » et « heures », la distribution se caractérise par une certaine concentration de la variance σ^2 et la longueur de la queue de distribution présente une décroissance en loi de puissance. Sur des fréquences plus basses, la concentration de

σ^2 est moins évidente et les queues de distributions transitent vers une décroissance exponentielle. Les distributions inverses gamma et log-normale sont toutes deux une distribution de loi de puissance et, comme le montre la figure 3.13, à mesure que les échelles de temps augmentent, les deux modèles deviennent s'adaptent moins aux queues de distribution empiriques.

Par conséquent, sur des échelles de temps à plus haute fréquence (« minute ») :

- l'analyse de l'exposant de la queue de distribution suggère que la distribution de la variance est corrélée à une distribution en loi de puissance.
- le test stationnaire KPSS ne peut rejeter l'hypothèse nulle de non-stationnarité suggérant une variance non constante dans le temps.
- la lente décroissance de la fonction d'autocorrélation ainsi que l'analyse de la distribution de la variance suggèrent que cette dernière présente une dépendance à long terme et un comportement plus complexe se caractérisant par une queue de distribution lourde.

Cette dernière analyse soutient le modèle multi fractal de Mandelbrot (1963) [78] pour simuler la fluctuation de la volatilité en utilisant une distribution stable de Lévy avec une moyenne espérée finie, mais une variance espérée infinie.

Deuxièmement, sur des échelles de temps plus grandes :

- la distribution de la variance devient moins corrélée à une distribution en loi de puissance à mesure que les échelles de temps augmentent.
- le test stationnaire ne peut rejeter l'hypothèse nulle sur les échelles de temps « heure » et « 4 heures » mais perd en significativité à mesure que les échelles de temps augmentent. Sur l'échelle de temps « journalière », le test de stationnarité rejette finalement l'hypothèse nulle de non-stationnarité.
- l'autocorrélation décroît de façon exponentielle et la queue de distribution devient moins lourde.

La suite de cette section se concentre sur la détermination de quelle super-statistique, entre la fonction inverse gamma et log-normal, est la plus significative pour décrire le dynamisme de la variance du Standard&Poor'500, selon différentes échelles de temps intra-journalières.

3.4.2 Calibration des modèles : Méthode « Metropolis-Hasting Sampler »

Avant de déterminer quel modèle de superstatistique est le plus approprié, nous devons préalablement, sur chaque échelle de temps, estimer les paramètres de chaque modèle qui permettent de simuler de manière significative les données observées. Nous souhaitons estimer les paramètres permettant de maximiser notre probabilité postérieure du paramètre fluctuant σ^2 conditionné aux données \mathcal{D}_n , comme décrit par l'équation (B.1, p.162).

Nous utilisons l'algorithme de marche aléatoire « Metropolis-Hasting » qui est une méthode spécifique de Markov Chain Monte Carlo (MCMC). Nous considérons l'équation (B.1, p.162) $p(\sigma^2|\mathcal{D}_n) \propto p(x|\mathcal{D}_n) * \pi(\sigma^2)$ comme notre distribution de proposition, avec $p(x|\mathcal{D}_n)$ la vraisemblance marginale et $\pi(\sigma^2)$ le préjugé. Ce dernier est proportionnel à la probabilité postérieure. La distribution du préjugé est définie sur un espace d'état χ qui est ergodique et stationnaire par rapport à π , ce qui signifie que si $\sigma^{2^t} \sim \pi(\sigma^2)$ alors $\sigma^{2^{t+1}} \sim \pi(\sigma^2)$ et, par conséquent, en construisant une chaîne de Markov sur χ , ce dernier aura une unique distribution stationnaire qui converge vers le préjugé π . La chaîne peut être considérée comme un échantillon de la distribution de proposition et le but est d'itérer des échantillons pour un paramètre donné σ^2 . Les critères d'acceptation de σ^2 dépendent du fait qu'en fonction de ce paramètre, la distribution du préjugé $\pi(\sigma^2)$ obtenue conjugué à la vraisemblance marginale doit rester proportionnelle à la distribution postérieure cible. En d'autres termes, σ^2 doit garantir que le préjugé reste proportionnel au postérieur.

Dans cette section, nous comparons deux modèles $\{M_1, M_2\}$:

- nous appelons $M_1 = \text{IGa}$ comme étant une distribution inverse gamma où :

$$g_1(\sigma^2; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (1/\sigma^2)^{\alpha+1} \exp(-\beta/\sigma^2) \quad (3.62),$$

avec $\sigma^2 > 0$ la volatilité financière, α le paramètre de forme, β le paramètre d'échelle et Γ la fonction gamma. Nous définissons la distribution du préjugé π sur σ^2 conditionnée à M_1 comme suit : $\pi(\sigma^2 | M_1) \sim g_1(\sigma^2; \alpha, \beta)$

- nous appelons $M_2 = \text{LN}$ comme étant une distribution log-normale où :

$$g_2(\sigma^2; \mu, s) = \frac{1}{s\sigma^2\sqrt{2\pi}} \exp\left(-\frac{\log^2(\sigma^2 - \mu)}{2s^2}\right) \quad (3.63),$$

avec $\sigma^2 > 0$ la volatilité financière, μ la moyenne et $s > 0$ le paramètre d'échelle. On définit la distribution du préjugé π sur σ^2 conditionnée à M_2 $\pi(\sigma^2 | M_2) \sim g_2(\sigma^2; \mu, s)$.

Pour exécuter l'algorithme de Metropolis-Hasting, nous procédons comme suit :

Algorithm. Metropolis-Hasting Random Walk

1. Generate prior \propto posterior distribution:

$$\pi(\sigma^2 | M_1) \sim p(\sigma^2 | \text{IGa}; \alpha, \beta)$$

$$\pi(\sigma^2 | M_2) \sim p(\sigma^2; s | \text{LN}; \mu, s)$$

2. Generate marginal likelihood distribution:

$$p(x | \sigma^2, M_{1,2}) \sim p(x | \text{Gaussian}, \mu = 0, \sigma^2)$$

3. Give two initial guess proposals σ_p^2 and $\sigma_p'^2$:

4. Calculate the ratio:

$$\rho_{M_1} = \frac{p(x | \sigma_p'^2, M_1) * \pi(\sigma_p'^2 | M_1)}{p(x | \sigma_p^2, M_1) * \pi(\sigma_p^2 | M_1)},$$

$$\rho_{M_2} = \frac{p(x | \sigma_p'^2, M_2) * \pi(\sigma_p'^2 | M_2)}{p(x | \sigma_p^2, M_2) * \pi(\sigma_p^2 | M_2)},$$

5. Accept or reject:

$$\text{if } \rho = \{\rho_{M_1}, \rho_{M_2}\} > 1: \text{ accept update parameter } \sigma_p'^2$$

$$\text{elif } \rho = \{\rho_{M_1}, \rho_{M_2}\} < 1: \text{ reject } \sigma_p'^2 \text{ and keep } \sigma_p^2$$

6. Update parameter:

If $\sigma_p'^2$ is accepted:

$$\sigma_{\text{update}}^2 = \sigma_p'^2 + \text{gradient} * \text{learning_rate} * \text{momentum}$$

else:

$$\sigma_{\text{update}}^2 = \sigma_p^2 + \text{gradient} * \text{learning_rate} * \text{momentum}$$

7. Repeat steps 4., 5., 6.

8. End after n iterations

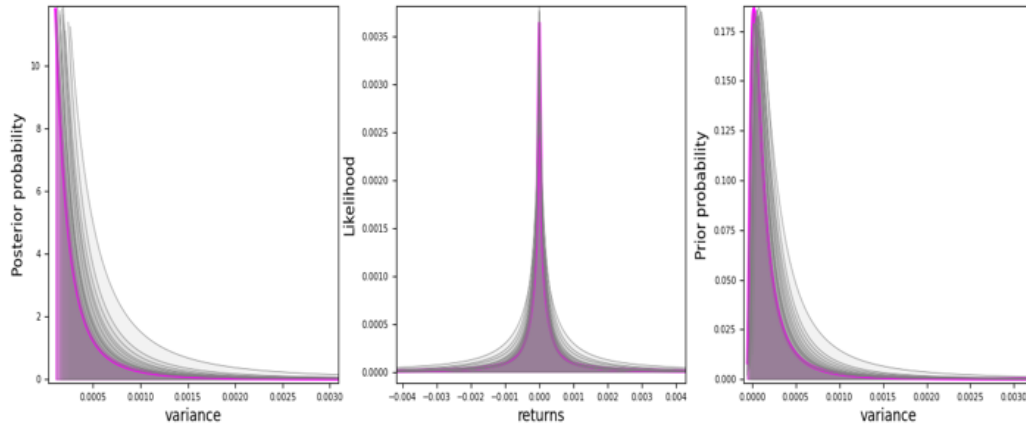


Figure 3. 14 Représentations graphiques des distributions de propositions issues des 15 meilleurs paramètres statistiques α, β du modèle $M_1 = \text{IGa}$ qui maximisent la probabilité postérieure. Ici, l'échelle de temps d'étude est « minute ». A gauche : probabilité postérieure. Milieu : probabilité marginale. A droite : probabilité a priori.

La couleur fuchsia représente les distributions de propositions issues des paramètres statistiques α, β les plus significatifs en termes de maximisation de la probabilité postérieure. Les couleurs grises représentent les 14 autres distributions de propositions issues des autres paramètres statistiques.

	$\text{IGa}(\sigma^2; \alpha, \beta)$	
Minute	$\alpha = 1.993$	$\beta = 0.00032$
Heure	$\alpha = 1.053$	$\beta = 0.00115$
4-heures	$\alpha = 0.951$	$\beta = 0.00222$
Journalière	$\alpha = 1.1196$	$\beta = 0.0063$

Tableau 3. 7 Estimation des paramètres du modèle IGa qui maximise la probabilité postérieure sur différentes échelles de temps.

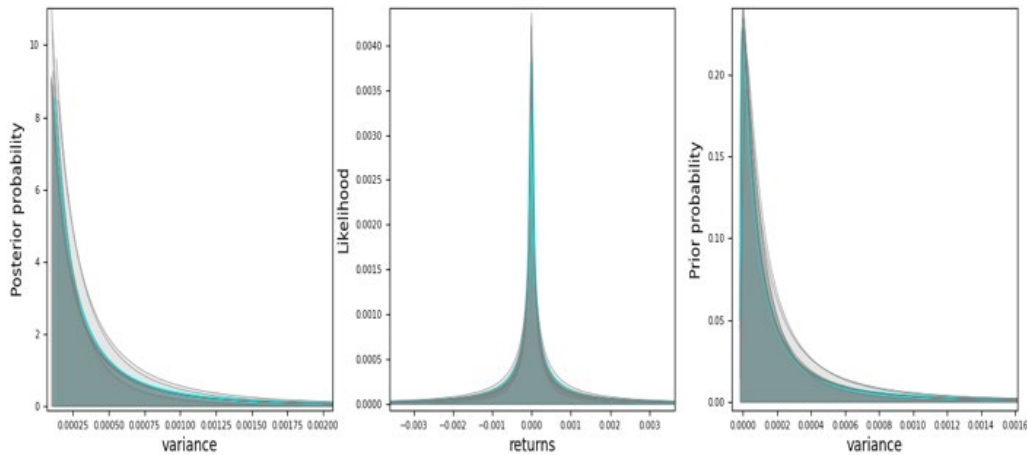


Figure 3. 15 Représentations graphiques des distributions de propositions issues des 15 meilleurs paramètres statistiques α, β du modèle $M_2 = LN$ qui maximisent la probabilité postérieure. Ici, l'échelle de temps d'étude est « minute ». A gauche : probabilité postérieure. Milieu : probabilité marginale. A droite : probabilité a priori. La couleur verte représente les distributions de propositions issues des paramètres statistiques α, β les plus significatifs en termes de maximisation de la probabilité postérieure. Les couleurs grises représentent les 14 autres distributions de propositions issues des autres paramètres statistiques.

	$LN(\sigma^2; \mu, s)$	
Minutes	$\mu = -1.69e - 05$	$s = 1.39$
Heure	$\mu = -6.51e - 05$	$s = 1.24$
4-heures	$\mu = -0.00019$	$s = 0.95$
Journalière	$\mu = -0.00083$	$s = 0.865$

Tableau 3. 8 Estimation des paramètres à partir de LN qui maximise la probabilité postérieure sur différentes échelles de temps.

Les figures 3.14 et 3.15 mettent en avant les résultats statistiques tirés des calibrations des modèles M1 et M2 (c'est-à-dire IGa, LN) via la méthode MCMC sur des échelles de temps en minutes. Les résultats statistiques sur d'autres échelles de temps (c.-à-d., « heure », « 4-heures », « journalière ») sont présentés par les figures C.1 et C.2 (p.164, 165) ainsi que les résultats statistiques de l'estimation des paramètres issus de la calibration des deux modèles (figure C.3, p.166). Les résultats soulignent comme attendu une certaine proportionnalité entre le préjugé et la probabilité postérieur. On constate également que la probabilité de vraisemblance marginale présente une queue de distribution plus épaisse qu'une distribution normale. L'estimation des paramètres maximisant la probabilité postérieure sur les deux

modèles sont présentées dans les tableaux 3.7 et 3.8. La section suivante compare la capacité des deux modèles à modéliser les données empiriques observées, en déterminant le facteur de Bayes sur les échelles de temps d'études respectives.

3.4.3 Comparaison de modèles via facteurs de Bayes

Nous considérons la série de données $\mathcal{D}_n = \{x_1, \dots, x_n\}$ comme étant l'évolution du log-prix du futur indice S&P500 de taille n en fonction de la longueur des échelles de temps. On considère les deux mêmes modèles paramétriques précédents M_1 et M_2 avec le préjugé sur leur probabilité respective, à savoir $\pi(\sigma^2|M_1)$ et $\pi(\sigma^2|M_2)$.

D'après le théorème de Bayes, la probabilité postérieure de M_1 et M_2 conditionnée aux données \mathcal{D}_n est :

$$\mathbb{P}(M_1, \sigma^2 | \mathcal{D}_n) = \frac{p(\mathcal{D}_n | M_1, \sigma^2) \pi(\sigma^2 | M_1)}{\sum_{k=1}^2 p(\mathcal{D}_n | M_k, \sigma^2) \pi(\sigma^2 | M_k)} \quad (3.64)$$

$$\mathbb{P}(M_2, \sigma^2 | \mathcal{D}_n) = \frac{p(\mathcal{D}_n | M_2, \sigma^2) \pi(\sigma^2 | M_2)}{\sum_{k=1}^2 p(\mathcal{D}_n | M_k, \sigma^2) \pi(\sigma^2 | M_k)} \quad (3.65),$$

où:

$$p(\mathcal{D}_n | M_1, \sigma^2) = \int \mathcal{L}_1(M_1, \sigma^2) \pi(\sigma^2 | M_1) d\sigma^2 \quad (3.66)$$

$$p(\mathcal{D}_n | M_2, \sigma^2) = \int \mathcal{L}_2(M_2, \sigma^2) \pi(\sigma^2 | M_2) d\sigma^2 \quad (3.67)$$

et $\{\mathcal{L}_1, \mathcal{L}_2\}$ sont la fonction de vraisemblance pour les modèles M_1 et M_2 .

Le facteur de Bayes entre les deux modèles est obtenu en divisant l'équation (3.66) par (3.67) :

$$\mathcal{BF}(\mathcal{D}_n) = \frac{\int \mathcal{L}_1(\sigma^2) \pi(\sigma^2) d\sigma^2}{\int \mathcal{L}_2(\sigma^2) \pi(\sigma^2) d\sigma^2} \quad (3.68)$$

L'utilisation du facteur de Bayes doit être pensée comme une alternative bayésienne

aux tests d'hypothèses classiques dans le choix de modèles. D'après l'équation (3.68), si pour une échelle de temps donnée, on a $\mathcal{BF}(\mathcal{D}_n) > 1$, alors le modèle M_1 doit être préféré à M_2 . La réciproque est également vraie.

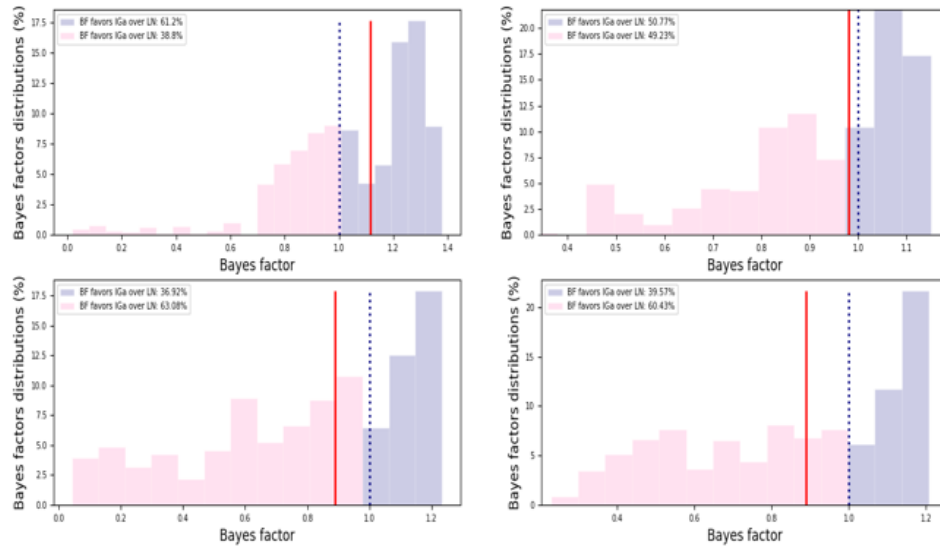


Figure 3.16 Distributions des facteurs de Bayes (BF) des séries de variables de variance simulées (c.-à-d. σ^2) conformément aux modèles estimés M_1 et M_2 . En haut à gauche : période = « minute ». En haut à droite : période = « heure ». En bas à gauche : période = « 4 heures ». En bas à droite : période = « journalière ». Les histogrammes bleus correspondent à $\mathcal{BF} > 1$ et les histogrammes roses à $\mathcal{BF} < 1$. La ligne verticale rouge correspond à la médiane des facteurs bayésiens, et la ligne bleue en pointillé est fixée à 1.

Nous proposons de déterminer une série de facteurs de Bayes au travers de la simulation d'un ensemble d'échantillons de probabilités postérieures de σ^2 issues du modèle M_1 et M_2 . Nous pensons qu'il est plus significatif de travailler dans un intervalle de confiance de facteurs de Bayes plutôt qu'avec une simple et unique valeur discrète issue d'une seule simulation.

La figure 3.16 représente les distributions statistiques des facteurs de Bayes (BF) obtenues à partir d'une série de variables simulées σ^2 conditionnées par les modèles estimés M_1 et M_2 et selon différentes échelles de temps. Nous constatons que, sur l'échelle « minute », 61,2 % des facteurs bayésiens simulés favorisent le modèle M_1 par rapport au modèle M_2 . De même, à mesure que les échelles de temps augmentent, nous constatons une transition des facteurs de Bayes favorisant le modèle M_2 avec plus

de 60% de facteurs bayésiens favorisant M_2 à M_1 . Cependant, pour chaque échelle de temps, la médiane des facteurs de Bayes est très proche de la valeur critique de 1 et reste toujours comprise dans l'intervalle $[0,33, 3]$. Par conséquent, il n'y a aucune preuve statistique claire qu'un modèle devrait être préféré à l'autre. En termes d'inférence Bayésienne, pour chaque échelle de temps, les deux modèles décrivent aussi bien la dynamique du paramètre fluctuant σ^2 .

3.4.4 Capacité des modèles *IGa et LN*¹⁴ à simuler les rendements logarithmiques

Nous avons utilisé le facteur de Bayes comme solution alternative aux tests d'hypothèses classiques dans le choix des modèles pour simuler la volatilité stochastique des rendements boursiers. Dans la section précédente, nous avons montré que les deux modèles simulent de manière équivalente la probabilité du préjugé $g(\sigma^2)$. Nous voulons maintenant déterminer le niveau de significativité des deux modèles quant à leur capacité à simuler les rendements logarithmiques empiriques \mathcal{D}_n du Standard&Poor500.

En nous appuyant sur l'équation bayésienne B.1 (p.162), nous considérons comme E. Van der Straeten et C. Beck [109], que les rendements boursiers sont le produit d'une dynamique lente liée à la fluctuation de la variable σ^2 et d'une dynamique rapide liée à la vitesse de relaxation du système vers un processus gaussien. Par conséquent, nous considérons ici que les rendements boursiers peuvent être modélisés par le produit de $N(0, 1) * IGa(\sigma^2; \alpha, \beta)$ ou par le produit de $N(0, 1) * LN(\sigma^2; \mu, s)$. Nous définissons l'hypothèse que les données empiriques \mathcal{D}_n n'ont pas la même distribution sous-jacente avec les données simulées comme étant l'hypothèse alternative H_1 , et nous définissons l'hypothèse nulle H_0 que les données observées et simulées partagent la même distribution.

¹⁴ Nous rappelons que IGa signifie inverse gamma et LN signifie log-normal.

p-value	Model 1 $N(0, 1) * IGa(\sigma^2; \alpha, \beta)$	Model 2 $N(0, 1) * LN(\sigma^2; \mu, s)$
Minute	0.0566 > 0.05	0.0129 < 0.05
Heure	1.26e-15 < 0.05	0.026 < 0.05
4-heures	1.89e-20 < 0.05	0.1144 > 0.05
Journalière	6.53e-06 < 0.05	0.064 > 0.05

Tableau 3. 9 Test p-value avec la méthode Kolmogorov-Smirnov (KS). Deux modèles (Modèle 1 & Modèle 2) sont testés par rapport aux données empiriques \mathcal{D}_n . Lorsque p-value > 0,05, nous ne rejetons pas l'hypothèse nulle.

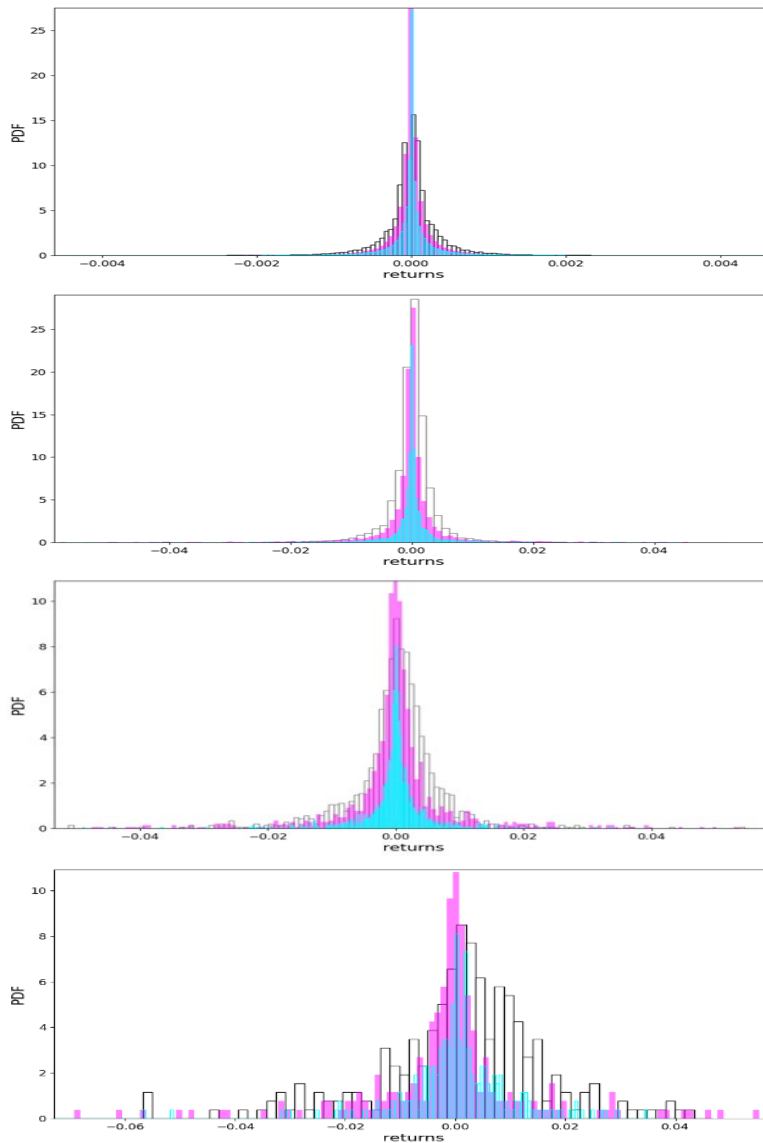


Figure 3. 17. Comparaison de la distribution de la densité de probabilité entre les données observées \mathcal{D}_n et les log-retours simulés du modèle 1 et du modèle 2. De haut en bas : la période va de « minute » à « journalière ». Histogrammes blancs avec bordure noire : distribution des données empiriques. Histogrammes fuchsia représente la distribution des données simulées issues du Modèle 1. Histogrammes cyan représente la distribution des données simulées issues du Modèle 2.

Le tableau 3.9 et la figure D.1 (p.167) présentent les résultats statistiques du test de p-value par la méthode Kolmogorov-Smirnov (KS) concernant l'estimation des rendements logarithmiques observés \mathcal{D}_n par le Modèle 1&2. Les paramètres utilisés pour calibrer les deux modèles sont ceux estimés dans la section précédente et résumés dans les tableaux 3.7 et 3.8. Nous montrons que, sur des échelles de temps « minute », le modèle 1 ne rejette pas l'hypothèse nulle à un niveau significatif de 5 % et permet de manière significative de modéliser la distribution des rendements boursiers observés. À mesure que les échelles de temps s'allongent, contrairement au modèle 1, le modèle 2 ne rejette pas l'hypothèse nulle et est donc préféré au modèle 1 pour l'estimation de la distribution des données \mathcal{D}_n .

Sur l'échelle de temps d'une heure, si les deux modèles rejettent l'hypothèse nulle, le modèle 2 reste le plus significatif.

Ce test de comparaison de modèles par p-value sur les rendements boursiers va de pair avec les résultats statistiques obtenus lors de la comparaison des modèles IGa et LN sur la distribution de la variance des données observées.

Des représentations graphiques de la distribution des rendements boursiers par rapport aux distributions résultant du modèle 1 et du modèle 2 sont illustrées par la figure 3.17. De l'échelle de temps « minute » à l'échelle de temps « 4 heures », les queues de distributions des rendements logarithmiques sont modélisées de manière significatives par les données simulées issus des modèles. Cependant, sur des échelles de temps « journalière », la forte concentration de données autour de la moyenne et la décroissance en loi de puissance de la distribution des données simulées ne représentent pas fidèlement les comportements des données observées.

Ce constat s'explique par l'analyse des données observées mise en avant dans la section 3.4.1. Nous montrons qu'à mesure où les échelles de temps augmentent, en particulier sur une échelle « journalière », la dynamique de la variance a tendance à être sans mémoire ; sa distribution ne présente plus de décroissance en loi de puissance et nous rejetons l'hypothèse nulle de non-stationnarité. Par conséquent, même si statistiquement parlant, la distribution log-normale modélise de manière significative le paramètre fluctuant de la série financière sur une base journalière, sur cette échelle de temps son utilisation risque de biaiser le comportement sous-jacent des données. En

effet, étant donné que les modèles inverses gamma et log-normal ont tous deux cette caractéristique de faire émerger des distributions avec de queues en loi de puissance, nous pouvons suggérer que, sur des périodes de plus basses fréquences, des études sur d'autres statistiques de la même famille sans décroissance de loi de puissance comme la distribution gamma doivent être effectuées. C'est que nous proposons de détailler dans la section suivante en donnant une preuve empirique de l'existence d'une transition de statistique de la complexité des séries financières, une transition de haute vers basse fréquence.

3.5 Preuve d'une transition de statistiques de la volatilité stochastique

Nous proposons d'analyser la fonction de distribution de volatilité $g(\sigma^2)$ du Standard&Poor500 du 01 janvier 2016 au 16 juillet 2021 selon quatre échelles de temps différentes à savoir « minute », « journalière », « hebdomadaire » et « mensuelle » et en fonction de trois distributions choisies comme préjugés : la distribution log-normal (LN), inverse-gamma (IGa) et gamma (Ga). Contrairement aux travaux de Tao Ma and R.A. Serota [76], nous décidons de ne pas traiter la loi inverse-gamma généralisée (GIGa) car, selon la valeur de ses paramètres β et γ , les distributions LN, IGa et Ga sont des cas particuliers de GIGa.

Comme précédemment, nous considérons le paramètre μ des rendements logarithmiques égal à zéro, et nous calculons la variance en élevant au carré les rendements logarithmiques. La figure 3.18 illustre la série temporelle utilisée pour cette étude sur la période « minute » (les graphiques sur les autres périodes sont présentés dans la figure E.1, p.168).

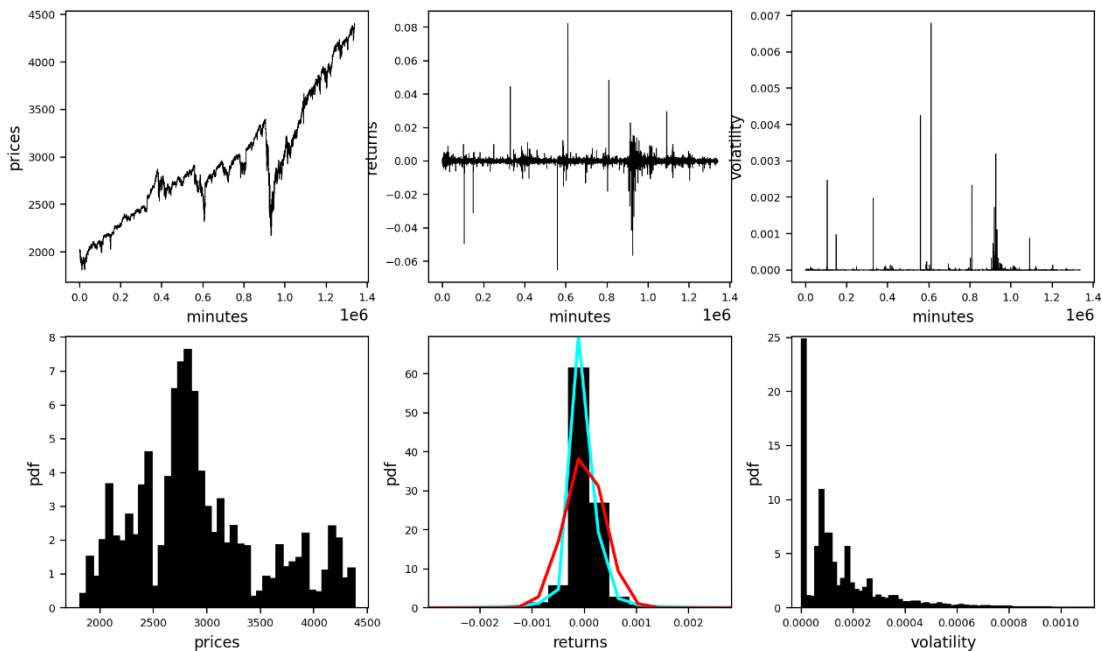


Figure 3. 18 Séries temporelles Standard&Poor500 sur l'échelle « minute » du 1er janvier 2016 au 16 juillet 2021.

En haut à gauche : les prix de Standard&Poor500. Milieu supérieur : rendements logarithmiques du Standard&Poor500. En haut à droite : rendements logarithmiques élevés au carré du Standard&Poor500. En bas à gauche : distribution des prix du Standard&Poor500. Milieu inférieur : distribution des rendements logarithmiques du Standard&Poor500. La courbe de couleur cyan est un ajustement de la distribution des données avec la loi t-Student. La courbe de couleur rouge correspond à une distribution normale. En bas à droite : distribution des rendements logarithmiques élevés au carré du Standard&Poor500.

Afin de calibrer les différents modèles (IGa, LN et Ga) et d'estimer leurs paramètres sur les différentes échelles de temps, nous proposons de minimiser la probabilité négative de leur vraisemblance respective au travers d'une optimisation bayésienne, en utilisant la méthode des processus gaussiens. L'estimation des paramètres des différents modèles sur l'échelle de temps « minute » est résumée dans le tableau 3.10 (les résultats pour les autres périodes sont indiqués dans le tableau F.1, p.169). La figure F.1 (p.170) compare la probabilité de chaque distribution selon différentes périodes. Nous constatons que la significativité de l'estimation des modèles augmente à mesure que la fréquence temporelle augmente.

Minutes Data	2016	2017	2018	2019	2020	2021
IGa(x, α , β)	α :1.98, β :4.18e-08	α :1.97, β :1.30e-08	α :1.97, β :3.25e-10	α :1.97, β :2.44e-08	α :0.468, β :3.52e-09	α :1.97, β :2.42e-08
LN(x, μ , σ)	μ : -5.40e-25 σ : 6.12	μ : -5.07e-24 σ :9.332	μ : -9.071e-28 σ : 6.624	μ : -1.49e-30 σ : 12.25	μ : -1.224e-22 σ : 10.61	μ : -8.22e-27 σ : 9.936
Ga(x, α , β)	α :2.96, β :0.793	α :0.00015, β :0.815	α :0.00033, β :0.815	α :7.096e-05, β :0.815	α :7.776e-05, β :0.793	α :0.000187, β :0.8148

Tableau 3. 10 Paramètres estimés par maximisation de la vraisemblance des différents modèles sur l'échelle temporelle « minute ».

Nous comparons ensuite la capacité de chaque modèle à simuler les rendements logarithmiques observés élevés au carré en utilisant la méthodologie de la somme des erreurs au carré (SSE). Le modèle de distribution qui correspond aux données simulées avec le SSE le plus faible est préféré. Les résultats pour l'année 2016 sont mis en avant dans le tableau 3.11 (les résultats pour les autres années sont présentés dans le tableau F.2, p.171). La figure 3.19 représente, pour l'année 2016 et sur les quatre échelles de temps d'études, la distribution du paramètre fluctuant σ^2 sous forme d'histogrammes et elle la met en relief avec la distribution qui permet de simuler ces données de manière significative parmi les modèles IGa, LN et Ga. Les valeurs de la p-value, permettant de rejeter ou pas l'hypothèse que les modèles d'études simulent de manière significative les données observées pour l'année 2016, sont recensées dans le tableau 3.11 (les résultats pour les autres années sont présentés par les figures F.2 et F.3, p.172,173). Au travers des données statistiques recueillies par les Tableaux 3.11, F.2 (p.171) et les Figures 3.19, F.2 et F.3 (p.172, 173), nous constatons sur l'ensemble des années d'études, une transition de statistiques d'une période de haute fréquence à une période de basse fréquence. En utilisant des données à haute fréquence (c'est-à-dire « minute »), la statistique log-normal (LN) est celle qui globalement permet de simuler le plus significativement les données observées. En utilisant des données de fréquence inférieure comme « journalière », nous constatons une mixture de statistiques avec, selon les années, le modèle log-normal (LN) ou inverse gamma (IGa) comme étant les deux les meilleurs candidats. Enfin, sur des données de basse fréquence à savoir « hebdomadaire » et « mensuelle », nous montrons une nette transition des statistiques vers une distribution gamma (Ga).

Year 2016	minutes	daily	weekly	monthly
Best fit distributions	LN(x,μ,σ)	IGa(x,α β)	LN(x,μ,σ)	Ga(x,α,β)
SSE from LN	1.57e+09	1.53e+08	5.93e+07	1.44e+08
SSE from IGa	1.64e+09	1.12e+08	1.38e+08	1.62e+08
SSE from Ga	1.63e+09	8.75e+08	6.29e+07	1.11e+08
p-value on best fit	0.173	0.0631	0.532	0.350

Tableau 3. 11 Comparaison des différents modèles avec la méthode de la somme des erreurs au carré pour l'année 2016 et selon différentes périodes. Le modèle de distribution qui correspond aux données simulées avec le SSE le plus faible est préféré. En 2016, le modèle LN fournit la meilleure estimation des données observées sur les échelles « minute » et « hebdomadaire ». Le modèle IGa fournit la meilleure estimation des données observées sur l'échelle « journalière » et le modèle Ga fournit la meilleure estimation des données observées sur l'échelle « hebdomadaire ».

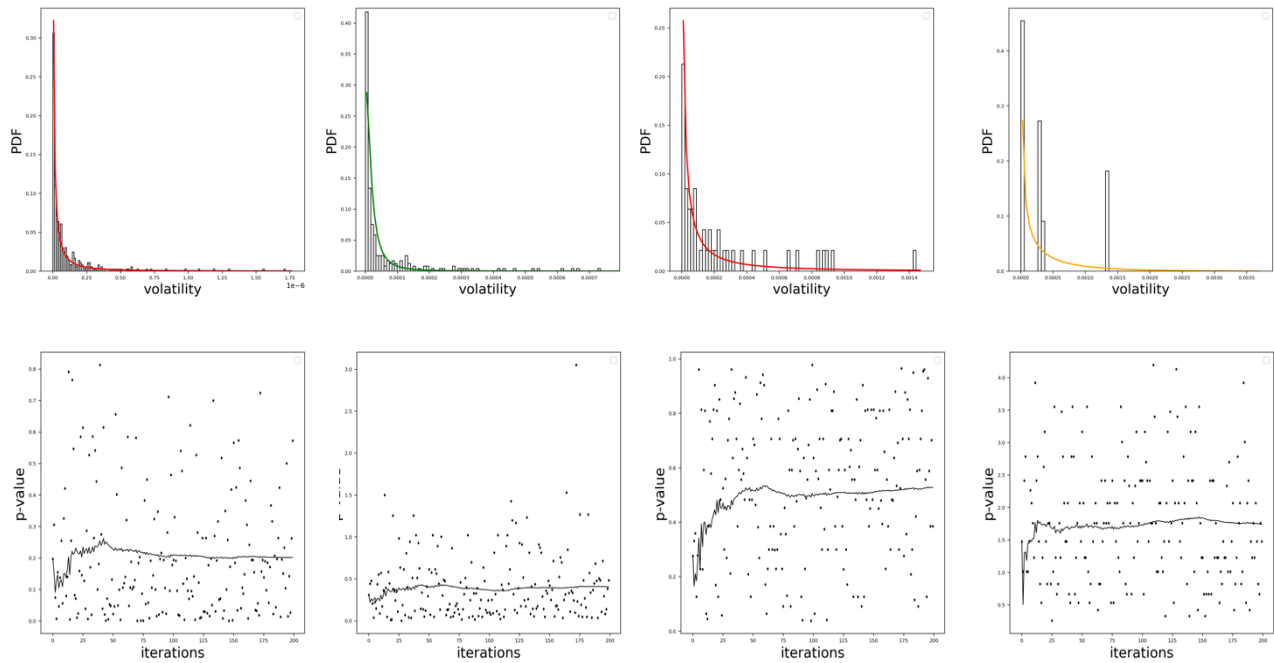


Figure 3. 19 Étude statistique pour l'année 2016. Sur le graphique supérieur et inférieur, de gauche à droite, l'échelle temporelle = « minute », « journalière », « hebdomadaire », « mensuelle ».

Graphiques supérieurs : comparaison de la distribution des données empiriques avec la distribution de données simulées issus du modèle le plus significatif. La courbe rouge correspond à la distribution des données simulées par le modèle LN, la courbe verte correspond à la distribution des données simulées par le modèle IGa et la courbe jaune correspond à la distribution des données simulées par le modèle Ga.

Graphiques inférieurs : Bootstrapping paramétrique de la p-value issu du test Kolmogorov-Smirnov concernant le modèle d'estimation le plus significatif parmi IGa, LN et Ga.

3.6 Conclusion

Dans ce Chapitre, nous partons de l'hypothèse que les rendements boursiers peuvent être considérés comme un macro-état présentant une dynamique spatio-temporelle inhomogène. Nous démontrons que la dynamique des cours boursiers résulte des interactions entre agents financiers définies par leur opportunité de profits respective. Nous avons supposé que la prise de décision d'un agent financier concernant le placement d'un pari repose sur la mise à jour de son opportunité de profits respective via l'inférence bayésienne et qu'elle est conditionnée aux résultats des paris précédents. Nous avons proposé des expériences numériques pour illustrer les comportements sous-jacents de cette dynamique en considérant les variations du cours des actions, à la fois dans un système conservateur et dans un système dissipatif.

Dans un système conservateur, nous avons démontré que l'opportunité globale de profits converge vers la probabilité marginale de profits d'un jeu équitable, c'est-à-dire 0,5, confirmant l'une des hypothèses d'Osborne. Nous mesurons également de légères asymétries d'opportunités de profits et avons démontré que, bien que nous ne puissions pas rejeter la distribution normale par p-value, la distribution t-Student est la distribution optimale pour s'adapter aux données simulées. Nous contredisons ainsi la théorie d'Osborne concernant la symétrie parfaite d'opportunités de profits entre agents et montrons que l'équilibre est atteint lorsque les données convergent vers une distribution normale.

Dans un système dissipatif, on considère le système comme un ensemble de « coarsed-grained » inhomogènes en interaction avec son environnement, ces grains représentant un groupe d'agents financiers. Les résultats des simulations numériques montrent que la dynamique du système peut être considérée comme une superposition de statistiques locales, toutes convergeant vers l'équilibre (c'est-à-dire convergeant vers une probabilité de profits de 0,5), mais en fonction de vitesses et de dynamiques différentes. Cette différence de vitesse de dynamique conduit à des asymétries dans la répartition des opportunités de profits du système et, à l'échelle globale, le système tend

vers l'équilibre sans l'atteindre. A partir des données simulées, nous montrons des preuves empiriques que la distribution des opportunités de profits est leptokurtique avec un nombre significatif de données loin de l'équilibre du système. Nous rejetons également par « p-value » l'hypothèse de normalité et montrons que la distribution t-Student correspond le mieux aux données observées. Nous rejetons la théorie d'Osborne concernant la symétrie parfaite des opportunités de profits entre les agents, rejetons la distribution normale ainsi que l'hypothèse de variance constante. En effet, la superposition des statistiques indique qu'à l'échelle globale, la dynamique du système est caractérisée par une volatilité fluctuante, une dynamique qui dépend de l'échelle spatio-temporelle d'étude.

Nous soutenons ensuite que la volatilité des actions doit être considérée comme une variable aléatoire avec une distribution $g(\sigma^2)$ et nous fournissons également des preuves théoriques que, via l'inférence bayésienne, en supposant que la vraisemblance des rendements logarithmiques a le temps de converger vers un processus gaussien, alors nous pouvons émettre le préjugé que la distribution $g(\sigma^2)$ est régie par loi inverse gamma.

Enfin, nous démontrons que d'autres distributions, à priori telles que la distribution gamma pour décrire le dynamisme de la variance, doivent être prises en considération en fonction de l'échelle de temps de l'étude. En effet, nous avons fourni des preuves empiriques que le dynamisme de la variance subit une transition de statistique selon que les données observées sont étudiées sur des échelles temporelles de hautes ou de basses fréquences. Sur des hautes fréquences (c'est-à-dire intra-journalières), les distributions log normales et inverses gamma se caractérisant par une loi de puissance permettent de manière équivalente de simuler significativement la distribution de σ^2 . Sur des données à basses fréquences de temps (c'est-à-dire « hebdomadaire » et « mensuelle »), nous observons une nette transition de la distribution des σ^2 vers une distribution gamma qui ne fait pas émerger de loi de puissance.

Cette dépendance du comportement de la variable σ^2 à l'échelle temporelle d'étude est importante car elle explique notamment le changement du comportement complexe des séries étudiées à l'échelle spatio-temporelle dans le Chapitre 2. Ce changement de complexité est à l'origine de certains phénomènes statistiques tels que la dépendance à long terme et le caractère non-markovien des séries financières que l'on trouve sur des hautes fréquences mais pas sur des basses fréquences. Enfin, ces résultats peuvent conduire à de meilleures approches de modélisation de la volatilité selon une période d'étude donnée et peuvent ouvrir la voie au développement de modèles alternatifs de volatilité.

Conclusion

Cette thèse de doctorat consacrée à l'étude de la dynamique des rendements logarithmiques des cours boursiers sous le prisme de la théorie des systèmes complexes nous a permis de mettre en exergue dans le chapitre 1 l'ensemble des caractéristiques associées à l'émergence d'une complexité de systèmes quelconques, afin d'en déduire les conditions nécessaires et suffisantes permettant d'en donner une définition la plus concise possible. Nous avons aussi montré l'importance qu'il y a de distinguer les systèmes dits chaotiques des systèmes complexes, de sorte à pouvoir développer des méthodes appropriées de mesures et modélisations de la complexité. Finalement, nous montrons que trois caractéristiques restent fondamentales lorsqu'on traite du sujet de la complexité, à savoir la nécessité que le système soit non statique, la capacité du système à s'auto organiser selon une adaptation locale et une capacité à s'adapter à son environnement de sorte que l'ensemble décrive un mouvement robuste et pérenne.

Pour donner suite à cette définition, nous en avons déduit que la complexité statistique était la méthode la plus rationnelle et fiable lorsque l'on souhaite aborder la question de la mesure. Nous avons ainsi étudié les rendements logarithmiques des cours boursiers selon différentes échelles spatio-temporelles en considérant le système comme étant un ensemble du type « coarse-grained », au travers de la Dynamique Symbolique et de la théorie de l'Information. Nous avons proposé notamment une analyse multi-fractale d'une série financière en générant les dimensions et entropies généralisées des séquences de symboles correspondantes. Nous cherchions dans ce chapitre 2 à répondre aux interrogations suivantes :

- Quelles propriétés émergent de l'évolution des rendements logarithmiques des cours boursiers ? ; ces propriétés se conservent-elles selon les échelles spatio-temporelles ?

- Sur quelles échelles spatio-temporelles, les rendements logarithmiques permettent-ils l'émergence de structures plus prévisibles ?

Pour répondre à ces questions, une analyse des rendements logarithmiques de l'indice Standard&Poor500 selon quatre échelles de temps intra-journalières, au travers de la méthode de la Symbolique Dynamique et de la théorie de l'Information, fut proposée. Nous avons mis en évidence plusieurs éléments :

- L'émergence de structures ordonnées et redondantes des probabilités conditionnelles des séquences de symboles sur des échelles de temps très courtes (« minute »), similaires aux formes issues des géométries fractales décrites par Mandelbrot.
- Un processus non-Markovien à l'origine du comportement des rendements, preuve d'une dépendance certaine entre les rendements futurs et les séries historiques, indiquant la présence de processus de mémoire.
- Sur des échelles petites (« minute »), l'occurrence d'événement futur nous transmet une information plus prévisible au sens de Shannon. Notre effet de surprise sur l'observation future de rendements est amoindri. Par conséquent, sur ce type d'échelle temporelle, nous aurons tendance à constater moins de désordre, moins d'incertitude et l'émergence de structure ordonnée plus prévisible.
- Le caractère multi-fractal et le processus de dépendance dans l'évolution des rendements logarithmiques ont aussi été mesurés par un spectre d'entropie généralisée de Rényi. Nous montrons numériquement l'évidence d'une certaine homogénéité et symétrie du spectre de singularité $f(\alpha(q))$ sur de petites échelles de temps (« minute ») et plus de variétés de valeurs de $f(\alpha(q))$ sur des échelles de temps plus grandes selon différentes valeurs de q avec une forte concentration de $f(\alpha(q))$ lorsque $q \rightarrow -\infty$.

- La valeur de l'entropie étant plus faible sur des échelles de temps plus petites, avec une décroissance monotone selon des variables de q , signifie qu'au sens de la théorie de l'Information, le comportement des rendements logarithmiques du Standard&Poor500 exhibe moins d'informations, par conséquent plus prévisibles. A contrario, sur des échelles de temps plus grandes, ce dernier devient plus imprévisible et exhibe plus d'informations au sens de Shannon.

Après avoir mis en évidence et mesuré le caractère complexe des rendements boursiers issus du Standard&Poor500, nous avons cherché au travers du chapitre 3 à répondre au problème de modélisation de ce phénomène complexe. L'objectif principal était de proposer une méthode alternative aux modèles standards de modélisations du processus de volatilité en étudiant le phénomène physique sous-jacents des séries financières. L'idée était de considérer la statistique de Gibbs-Maxwell-Boltzmann [53,58,83] concernant la modélisation du comportement d'un système thermodynamique hors équilibre et de montrer qu'une analogie forte pouvait être établie avec la dynamique des rendements boursiers. Dans cette démarche d'analogie et au travers de simulations numériques et d'inférences bayésiennes, nous avons démontré qu'il était possible de modifier les hypothèses de la théorie des mouvements boursiers d'Osborne et avons montré que :

- les rendements boursiers peuvent être considérés comme un macro-état présentant une dynamique spatio-temporelle inhomogène. Ils constituent un ensemble de macro-états locaux dits « coarse-grained », chacun en équilibre local (ou tendant vers l'équilibre local), mais défini selon des distributions de probabilités postérieures d'opportunités de variances différentes.
- la cause de cette inhomogénéité vient du caractère dissipatif des rendements boursiers.
- localement, les rendements boursiers sont décrits par une mécanique statistique d'équilibre avec des variables statistiques conservées, alors que, globalement, il existe encore d'autres statistiques avec des variables persistantes qui peuvent

être efficacement décrites par une superposition de plusieurs statistiques sur différentes échelles de temps et d'espaces.

- En raison d'un système non conservatif, la dispersion des rendements boursiers se définit par $\sigma_i\sqrt{\tau}$ avec σ_i la volatilité du marché étudié, que l'on considère comme étant une variable stochastique décrite par une distribution donnée selon un ou plusieurs paramètres de conservation.

Nous en avons déduit que ce résultat rejoignait une branche de la théorie de la mécanique statistique, à savoir la super-statistique. Nous avons ensuite cherché à modéliser le comportement de la variation absolue (ou élevée au carré) des rendements logarithmiques que l'on a associé à la complexité du système étudié. Le but étant de définir un modèle d'estimation de volatilité suivant différentes échelles de temps. Nous cherchions notamment à répondre aux questions suivantes :

- Les super-statistiques permettent-elles d'exprimer le processus statistique sous-jacent aux cours boursiers à l'origine de l'émergence de complexités se caractérisant entre autres par un processus de mémoire ?
- Quelle super-statistique parmi la famille des distributions généralisées permet en termes significatifs de modéliser la distribution de la variance des séries financières et ce, selon différentes échelles temporelles ?

Pour répondre à ces questions, nous avons proposé de modéliser le processus de variance de séries financières suivant deux super-statistiques particulières et selon différentes échelles temporelles intra-journalières. Le but est de déterminer si ces statistiques étaient significatives et si une l'était plus que l'autre selon l'échelle temporelle d'étude. Nous avons donc proposé par inférence Bayésienne de modéliser le processus de la variance au travers de la statistique inverse gamma et log-normal. Puis, nous avons élargi l'étude en modélisant la distribution des rendements logarithmiques au travers du produit de la probabilité de vraisemblance des rendements avec la distribution de la variance. Les résultats nous ont permis de conclure:

- Sur des échelles de temps relativement petites, en particulier sur l'échelle "minute", l'analyse des exposants en loi de puissance suggère que la distribution de la variance est corrélée à une distribution en loi de puissance. Le test stationnaire KPSS suggère la présence d'une variance non constante dans le temps. La décroissance de la fonction d'autocorrélation et l'analyse de l'histogramme de la distribution permettent de conclure que, sur des échelles de hautes fréquences, la variable de variance du Standard&Poor500 présente une dépendance à long terme et une dynamique complexe avec une queue de distribution lourde.
- A mesure que les échelles de temps augmentent, notamment en ce qui concerne l'échelle "journalière", nous observons une transition de statistique. La distribution de probabilité de la variance devient non corrélée à une distribution de loi de puissance ; la queue de distribution n'est plus épaisse et l'hypothèse nulle de non-stationnarité a été rejetée.
- Sur des échelles intra-journalières, au travers du ratio de Bayes, nous montrons que la modélisation du processus de variance non constante par la distribution inverse gamma et log-normal est équivalente. En revanche, en ce qui concerne la modélisation des rendements logarithmiques, nous avons montré, que sur l'échelle « minute », les rendements logarithmiques étaient mieux ajustés par la distribution du produit du modèle inverse gamma et normal, tandis que, pour des échelles intra-journalières plus grandes, le produit du modèle log-normal avec la distribution normale était un meilleur candidat.
- A partir de l'échelle « journalière », si la statistique log-normal reste statistiquement significative pour modéliser le processus de la variance, nous observons tout de même une transition de statistique avec un processus sous-jacent devenant sans mémoire.

Par conséquent, afin de tenir compte de ce changement de statistique, nous avons proposé d'étudier des séries temporelles du Standard&Poor500 sur des échelles de temps de fréquences beaucoup plus basses et en tenant compte d'une autre distribution de la même famille de super-statistique, à savoir la distribution gamma. Finalement nous donnons des preuves empiriques que :

- Le dynamisme de la variance subit une transition de statistique selon que les données observées sont étudiées sur des échelles temporelles de hautes ou de basses fréquences. Sur des hautes fréquences (c'est-à-dire « intra-journalières »), les distributions log-normal et inverses gamma se caractérisant par une loi de puissance permettent de manière équivalente de simuler significativement la distribution de la variance. Sur des données à basses fréquences (c'est-à-dire « hebdomadaire » et « mensuelle »), nous observons une nette transition de la distribution du processus de la variance vers une distribution gamma. Nous notons qu'à l'opposé des distributions inverses gamma et log-normal, cette dernière ne fait pas émerger de loi de puissance.

Finalement, cette dépendance du comportement de la variance à l'échelle temporelle est importante et est à mettre en relation avec le changement du comportement complexe des séries étudié et mis en évidence dans le chapitre 2. Ce changement de complexité est à l'origine de certains phénomènes statistiques tels que la dépendance à long terme et le caractère non-markovien des séries financières.

L'introduction de la Dynamique Symbolique et la Théorie de l'information a permis une compréhension plus approfondie des caractéristiques des rendements d'un cours boursier. Elle a notamment permis le développement d'un raisonnement aboutissant à la construction d'une méthode de modélisation des cours boursiers au travers d'un modèle de volatilité s'appuyant sur une branche de la mécanique statistique. Ce type de modèle permet une modélisation fidèle du comportement statistique sous-jacent au système étudié et indique l'échelle temporelle adéquate à la modélisation des caractéristiques souhaitées. Il fournit également une précision

supplémentaire, non seulement pour les modèles d'estimation de risque, puisqu'il prend en compte le potentiel caractère asymétrique des rendements en s'adaptant aux échelles temporelles d'études, mais également dans le développement de stratégie financière d'investissement telle que des modèles dits de « trading de flux ».

Enfin, ces résultats peuvent conduire à de meilleures approches de modélisation de la volatilité selon une période d'étude donnée et peuvent ouvrir la voie au développement de modèles alternatifs de volatilité stochastique. Selon le type de données observées que l'on souhaite modéliser, nous avons montré que l'un des avantages de la statistique généralisée de Boltzmann est qu'elle contient plusieurs familles de distributions exponentielles qui se définissent toutes par des caractéristiques particulières permettant ainsi de s'adapter à des données dont la dynamique évolue selon la taille de l'échantillon ou encore selon l'échelle temporelle. Par exemple, dans le cas où la statistique de Boltzmann converge vers une distribution inverse gamma ou log normal, alors cette dernière permettra de modéliser de manière significative des volatilités relativement larges. De la même manière, si la statistique de Boltzmann converge vers une distribution gamma, alors cette dernière permettra de modéliser des volatilités relativement faibles. Cette capacité d'auto-ajustement selon les données observées est un atout non négligeable au regard par exemple des modèles multiplicatifs ¹⁵qui ont le désavantage de se définir uniquement par une distribution qui décroît de manière exponentielle pour des valeurs de volatilités faibles, ou encore le modèle Heston ¹⁶dont la distribution de la volatilité converge vers une distribution gamma ne permettant pas d'avoir une estimation fiable pour des valeurs larges de volatilités. Enfin, contrairement à ces deux types de modèles, la statistique généralisée de Boltzmann a l'avantage d'être facilement calibrée sans coûts de calculs ou de paramétrisation excessifs. Dans le monde de la finance à haute fréquence (intra-journalière), cet élément est loin d'être négligeable.

¹⁵ Gerig, J. Vicente, M. A. Fuentes, Model for non-gaussian intraday stock returns, *Physical Review E* 80 (2009) 065102R.

¹⁶ Heston, S. L. (1993). A closed-form solution for options with stochastic volatility, with applications to bond and currency options. *Review of Financial Studies* 6, 327–43.

Annexe A

Illustrations des expériences numériques et résultats statistiques

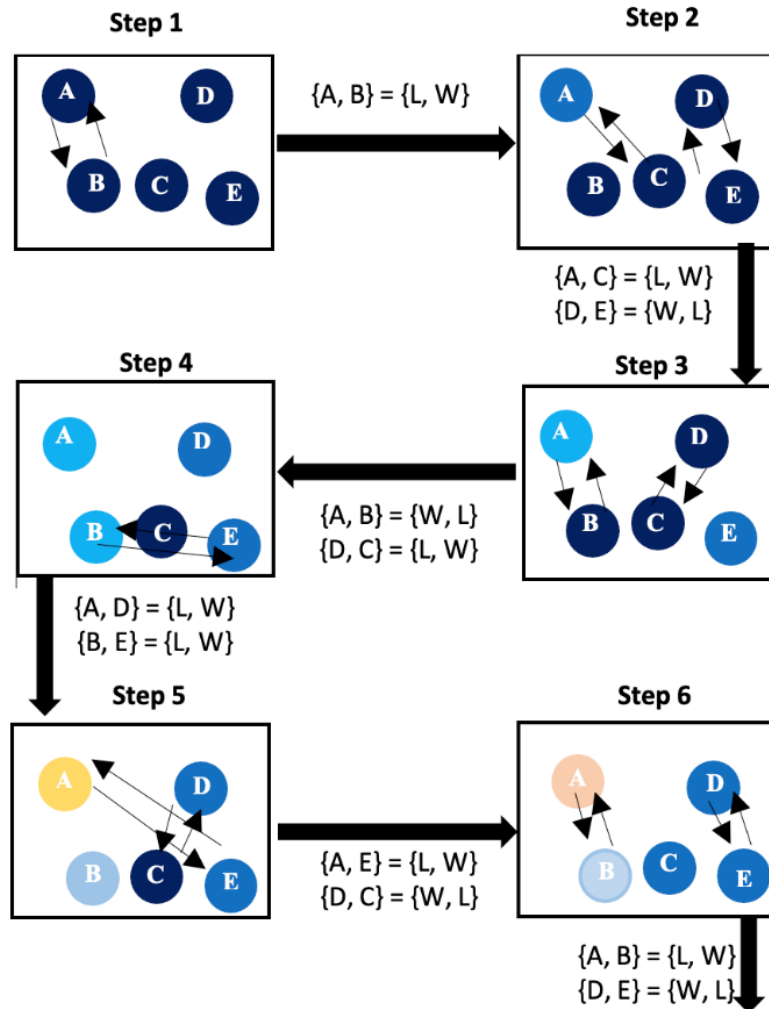


Figure A. 1 Illustration simplifiée de la simulation numérique dans le cas d'un système conservatif. A chaque étape au moins un pari intervient entre deux micro-états. Exemple : à l'étape 1, A parie contre B, A perd et B gagne. Puis, A et B mettent à jour leur probabilité de profits de manière bayésienne.

Step 1	A	B	C	D	E	
Win	1	1	1	1	1	5
loss	0	0	0	0	0	0
$P(\text{win} \mathcal{M}_i)$	1	1	1	1	1	1
Step 2	A	B	C	D	E	
Win	1	2	1	1	1	6
loss	1	0	0	0	0	1
$P(\text{win} \mathcal{M}_i)$	0.5	1	1	1	1	0.9
Step 3	A	B	C	D	E	
Win	1	2	2	2	1	8
loss	2	0	0	0	1	3
$P(\text{win} \mathcal{M}_i)$	0.16	1	1	1	0.27	0.69
Step 4	A	B	C	D	E	
Win	2	2	3	2	1	10
loss	2	1	0	1	1	5
$P(\text{win} \mathcal{M}_i)$	0.33	0.5	1	0.5	0.33	0.53
Step 5	A	B	C	D	E	
Win	2	2	3	3	2	12
loss	3	2	0	1	1	7
$P(\text{win} \mathcal{M}_i)$	0.28	0.37	1	0.6	0.54	0.56
Step 6	A	B	C	D	E	
Win	2	2	3	4	3	14
loss	4	2	1	1	1	9
$P(\text{win} \mathcal{M}_i)$	0.24	0.39	0.66	0.72	0.66	0.53
Step 7	A	B	C	D	E	
Win	2	3	3	5	3	16
loss	5	2	1	1	2	11
$P(\text{win} \mathcal{M}_i)$	0.22	0.51	0.67	0.77	0.51	0.54

Tableau A. 1 Évolution de la probabilité postérieure de chaque micro-état et à mesure que les paris progressent dans le temps.

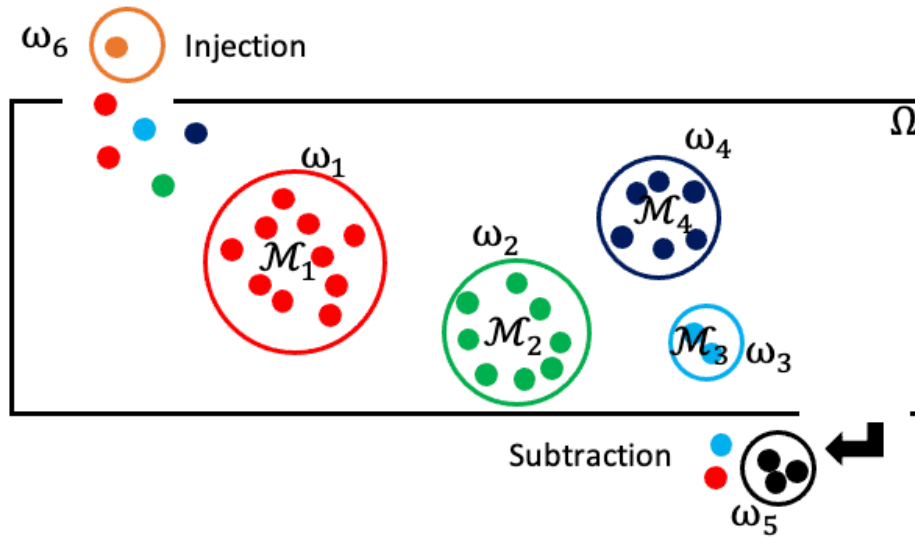


Figure A. 2 Illustration simplifiée de la simulation numérique dans le cas d'un système dissipatif. Macro-état dissipatif Ω constitué de macro-états locaux du types « coarse-grained » ($\omega_1, \omega_2, \omega_3 \dots \omega_N$) qui interagissant entre eux.

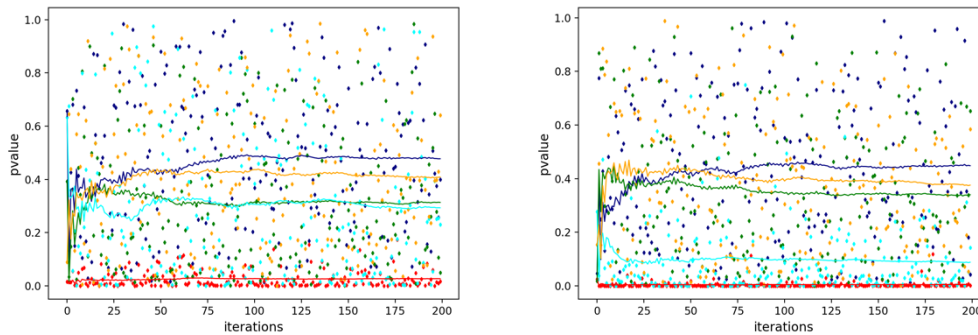


Figure A. 3 Comparaison de la distribution des données observées du macro-état Ω issu des deux expériences (conservatif et dissipatif) avec des données simulées par cinq distributions (t-Student, Laplace, Cauchy, Pearson et Normal).

Graphique de Gauche : données issues du système conservatif.

Graphique de droite : données issues du système dissipatif.

Résultats du test Kolmogorov-Smirnov comme test statistique de significativité.

L'hypothèse nulle est que les données simulées modélisent de manière significative les données observées. Si $p\text{-value} > 0,05$ alors nous ne rejetons pas l'hypothèse nulle.

Une série de p-value est calculée via la méthode de bootstrapping paramétrique

(Bleu : p-value issue du test statistique sur le modèle t-Student, jaune : p-value issue du test statistique sur le modèle Laplace, vert : p-value issue du test statistique sur le modèle Cauchy, rouge p-value issue du test statistique sur le modèle normale, cyan : p-value issue du test statistique sur le modèle Pearson).

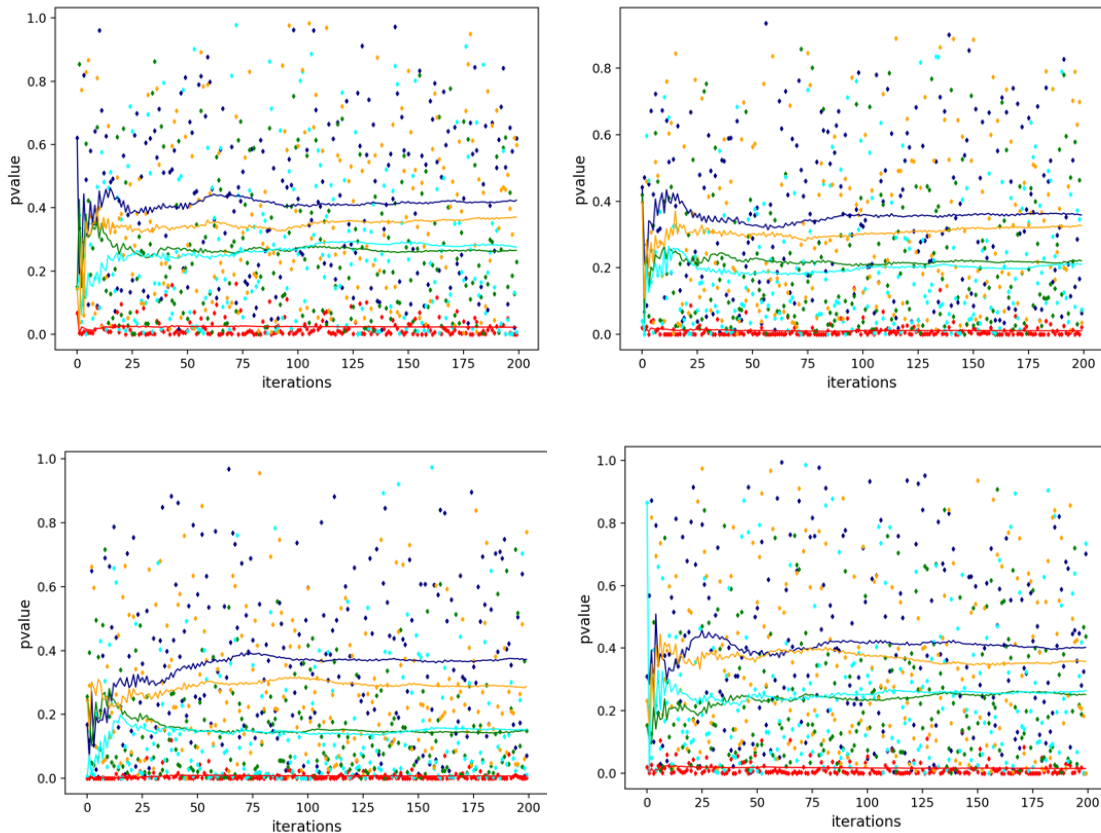


Figure A. 4 Comparaison de la distribution des données observées de chaque macro-état local ($\omega_1, \omega_2, \omega_3, \omega_4$) issu de l'expérience dans le cas d'un système dissipatif avec des données simulées par cinq distributions (t-Student, Laplace, Cauchy, Pearson et Normal).

Résultats du test Kolmogorov-Smirnov comme test statistique de significativité.

L'hypothèse nulle est que les données simulées modélisent de manière significative les données observées. Si $p\text{-value} > 0,05$ alors nous ne rejetons pas l'hypothèse nulle.

Une série de p-value est calculée via la méthode de bootstrapping paramétrique.

Graphique supérieur gauche : p-value issue du test statistique des modèles testés avec les données du « coarse-grained » ω_1 . Graphique supérieur droite : p-value issue du test statistique des modèles testés avec les données du « coarse-grained » ω_2 . Graphique inférieur gauche : p-value issue du test statistique des modèles testés avec les données du « coarse-grained » ω_3 . Graphique inférieur droite : p-value issue du test statistique des modèles testés avec les données du « coarse-grained » ω_4 ; (bleu : p-value issue du test statistique sur le modèle t-Student, jaune : p-value issue du test statistique sur le modèle Laplace, vert : p-value issue du test statistique sur le modèle Cauchy, rouge p-value issue du test statistique sur le modèle normale, cyan : p-value issue du test statistique sur le modèle Pearson).

Annexe B

Probabilité conjuguée de la distribution Normale [117]

Soient X_1, \dots, X_n n observations échantillonnées à partir d'une densité de probabilité $p(X|\sigma^2)$ avec σ^2 une variable aléatoire et $p(X|\sigma^2)$ représente la densité de probabilité conditionnelle de X conditionnée à σ^2 . En revanche, nous écrivons $p_{\sigma^2}(X)$ si nous considérons σ^2 comme une valeur déterministe.

D'après le théorème de Bayes, la distribution a posteriori peut s'écrire :

$$p(\sigma^2|X_1, \dots, X_n) = \frac{p(X_1, \dots, X_n|\sigma^2)\pi(\sigma^2)}{p(X_1, \dots, X_n)} = \frac{\mathcal{L}_n(\sigma^2)\pi(\sigma^2)}{c_n} \propto \mathcal{L}_n(\sigma^2)\pi(\sigma^2) \quad (\text{B.1})$$

Où $\mathcal{L}_n(\sigma^2) = \prod_{i=1}^n p(X_i|\sigma^2)$ est la fonction de vraisemblance et :

$c_n = p(X_1, \dots, X_n) = \int p(X_1, \dots, X_n|\sigma^2) \pi(\sigma^2) d\sigma^2 = \int \mathcal{L}_n(\sigma^2)\pi(\sigma^2) d\sigma^2$ est une constante normalisée que l'on appelle aussi évidence.

La distribution du préjugé est conjuguée si elle est fermée sous échantillonnage. Autrement dit, si P est une famille de distributions que l'on considère comme étant le préjugé et pour chaque σ^2 nous avons une distribution $p(\cdot|\sigma^2) \in \mathcal{F}$ sur échantillon de l'espace χ , alors si la probabilité postérieure :

$$p(\sigma^2|x) = \frac{p(x|\sigma^2) \pi(\sigma^2)}{\int p(x|\sigma^2) \pi(\sigma^2) d\sigma^2} \quad (\text{B.2}),$$

satisfait $p(\cdot|\sigma^2) \in P$, alors nous disons que la famille définie par P est conjuguée à la famille de distributions d'échantillonnage \mathcal{F} . Pour que cette notion soit significative, la famille P doit être suffisamment restreinte et est généralement considérée comme une famille paramétrique spécifique.

Nous pouvons notamment caractériser ces probabilités conjuguées pour les modèles de familles exponentielles. Supposons que $p(\cdot | \sigma^2)$ est un modèle gaussien avec μ connu, de sorte que le paramètre libre est la variance σ^2 . La fonction de vraisemblance est :

$$\begin{aligned} p(x_1, \dots, x_n | \sigma^2) &\propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2\right) \\ &= (\sigma^2)^{-n/2} \exp\left(-\frac{n}{2\sigma^2} \overline{(X - \mu)^2}\right) \quad (\text{B.3}) \end{aligned}$$

où,

$$\overline{(X - \mu)^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2. \quad (\text{B.4})$$

Dans ce cas particulier, la probabilité du préjugé conjuguée est une distribution inverse gamma. Rappelons que σ^2 a une distribution gamma inverse avec les paramètres α et β dans le cas $1/\sigma^2 \sim \text{Gamma}(\alpha, \beta)$; la densité prend la forme :

$$\pi_{\alpha, \beta}(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} e^{-\beta/\sigma^2} \quad (\text{B.5}).$$

En considérant cette loi correspondre à la distribution du préjugé, la distribution de la probabilité postérieure du paramètre libre σ^2 est donnée par :

$$\sigma^2 | X_1, \dots, X_n \sim \text{InvGamma}\left(\alpha + \frac{n}{2}, \beta + \frac{n}{2} \overline{(X - \mu)^2}\right) \quad (\text{B.6}).$$

Annexe C

Résultats statistiques issus de la maximisation des probabilités postérieures via MCMC

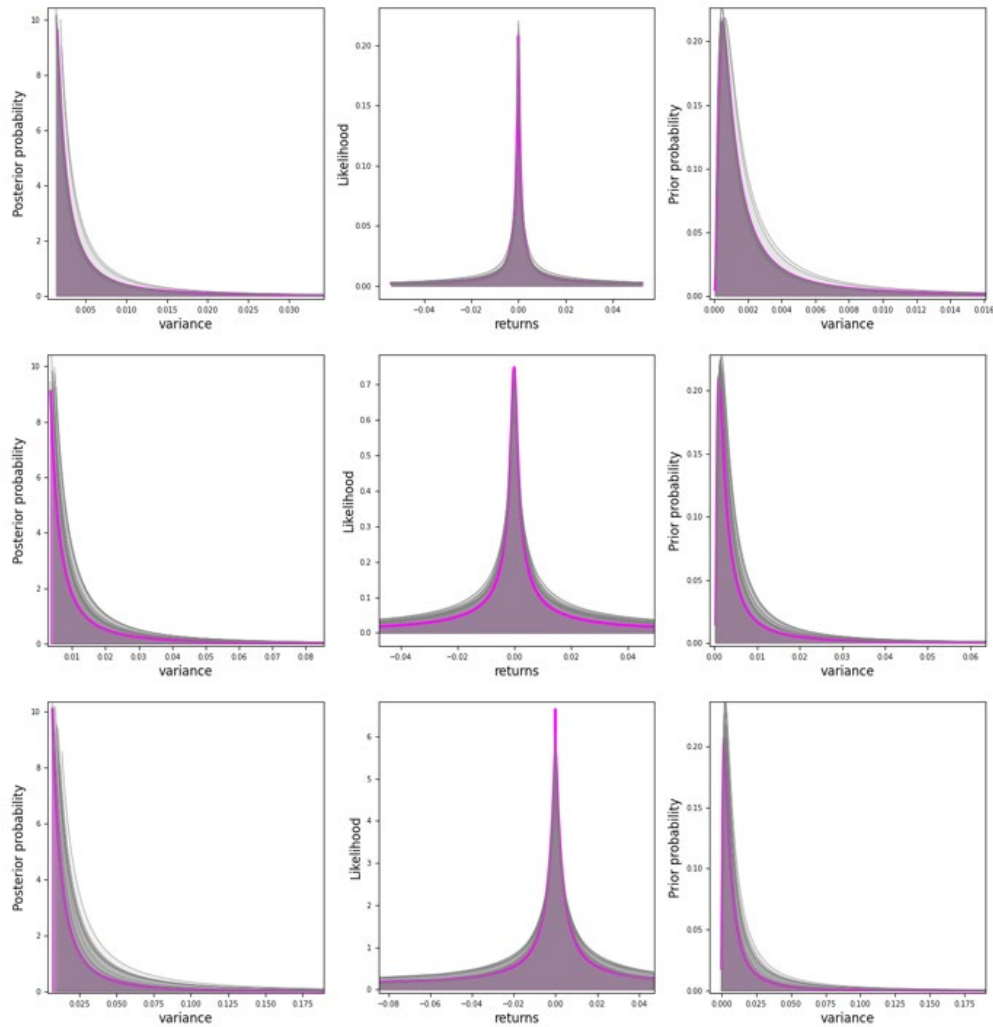


Figure C. 1 Représentations graphiques des distributions de propositions issues des 15 meilleurs paramètres statistiques α, β du modèle $M_1 = \text{IGa}$ qui maximisent la probabilité postérieure. Ici, l'échelle de temps d'étude partant du graphique de haut en bas est : « heure », « 4heures » et « journalière ». A gauche : probabilité postérieure. Milieu : probabilité marginale. A droite : probabilité a priori. La couleur fuchsia représente les distributions de propositions issues des paramètres statistiques α, β les plus significatifs en termes de maximisation de la probabilité postérieure. Les couleurs grises représentent les 14 autres distributions de propositions issues des autres paramètres statistiques.

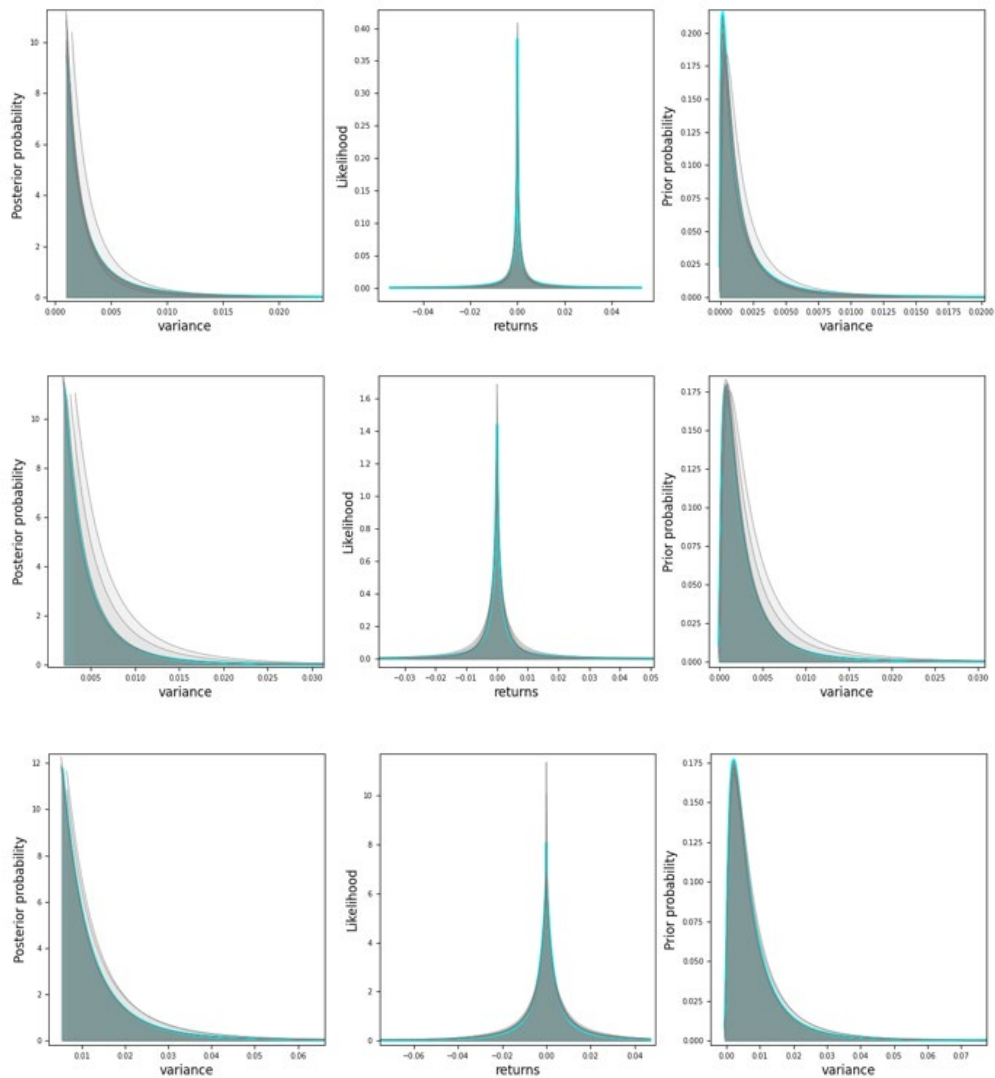


Figure C. 2 Représentations graphiques des distributions de propositions issues des 15 meilleurs paramètres statistiques α, β du modèle $M_2 = LN$ qui maximisent la probabilité postérieure. Ici, l'échelle de temps d'étude partant du graphique de haut en bas est : « heure », « 4heures » et « journalière ». A gauche : probabilité postérieure. Milieu : probabilité marginale. A droite : probabilité a priori. La couleur verte représente les distributions de propositions issues des paramètres statistiques α, β les plus significatifs en termes de maximisation de la probabilité postérieure. Les couleurs grises représentent les 14 autres distributions de propositions issues des autres paramètres statistiques.

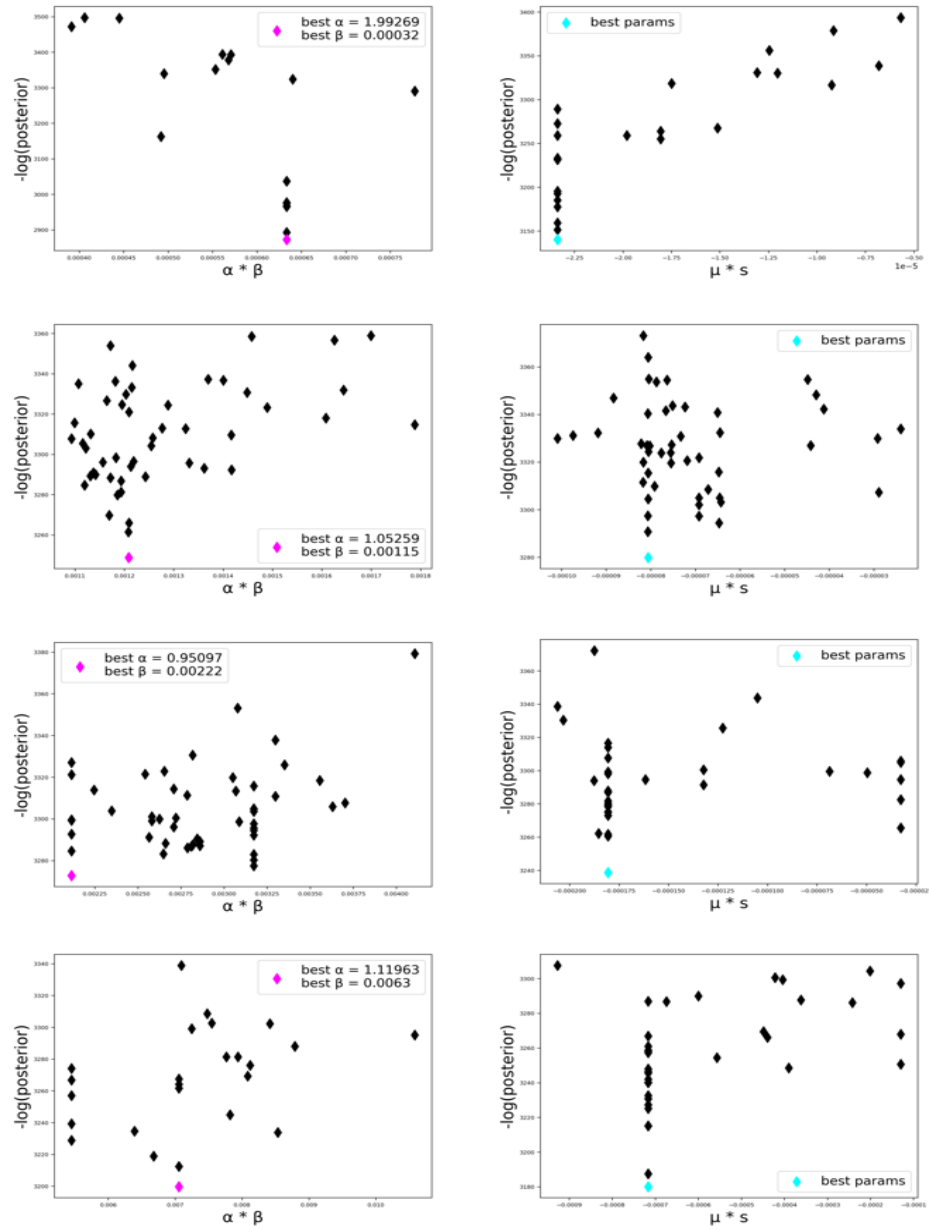


Figure C. 3 Résultats statistiques de l'estimation des paramètres issus de la calibration des deux modèles en minimisant la valeur négative de la probabilité postérieure logarithmique. Colonne de gauche : calibration du modèle $M_1 = \text{IGa}(\sigma^2; \alpha, \beta)$. Colonne de droite : calibration du modèle $M_2 = \text{LN}(\sigma^2; \mu, s)$. De haut en bas : l'échelle temporelle va de « minute » à « journalière ». Le diamant fuchsia représente la meilleure estimation des paramètres α et β . Le diamant cyan représente la meilleure estimation des paramètres μ et s .

Annexe D

Kolmogorov-Smirnov (KS) p-value test

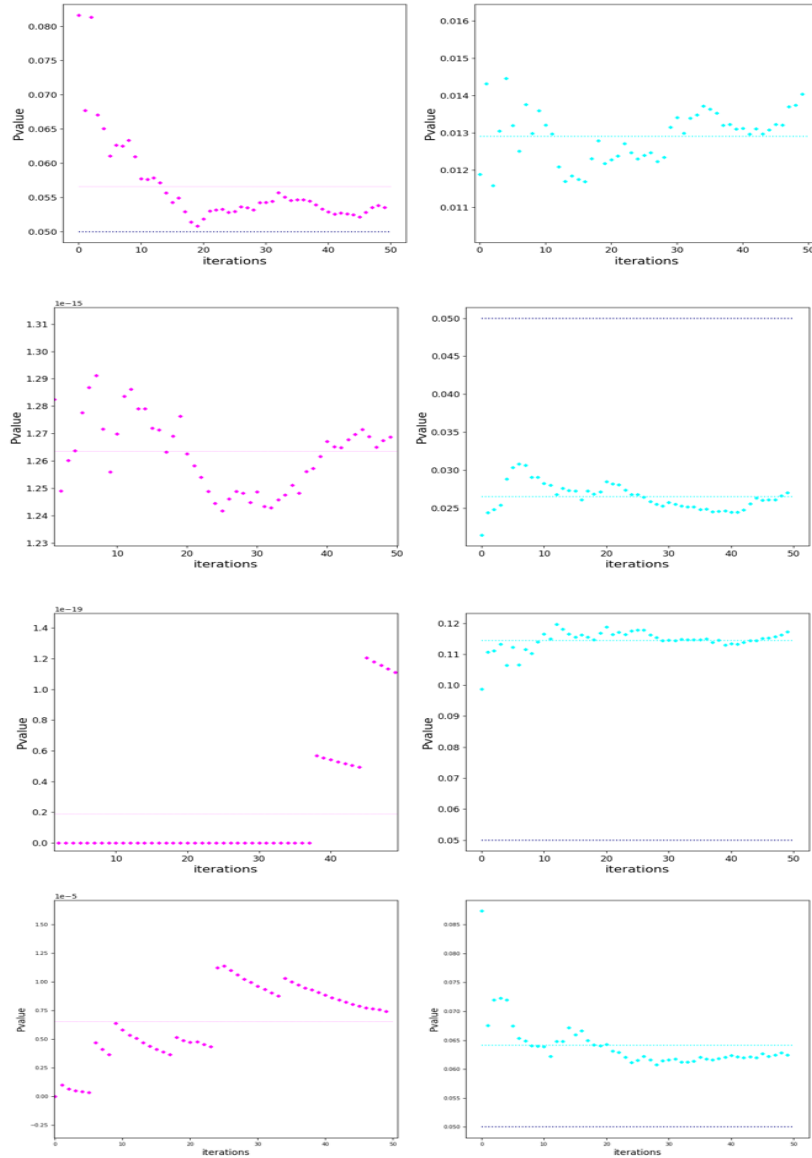


Figure D. 1 Résultats statistiques du test de p-value par la méthode Kolmogorov-Smirnov (KS) concernant l'estimation des rendements logarithmiques observés \mathcal{D}_n par le Modèle 1&2. Colonne de gauche : test statistique du modèle 1 par rapport aux rendements boursiers. Colonne de droite : test statistique du modèle 2 par rapport aux rendements boursiers. De haut en bas : l'échelle temporelle va de « minute » à « journalière ». La ligne pointillée horizontale fuchsia est la valeur moyenne des p-values du modèle 1 après 150 itérations. La ligne pointillée cyan horizontale est la valeur moyenne p-values du modèle 2 après 150 itérations. La ligne en pointillé bleu marine est configurée à la valeur de p-value = 0,05.

Annexe E

Séries temporelles du Standard&Poor500 étudiées selon différentes échelles temporelles

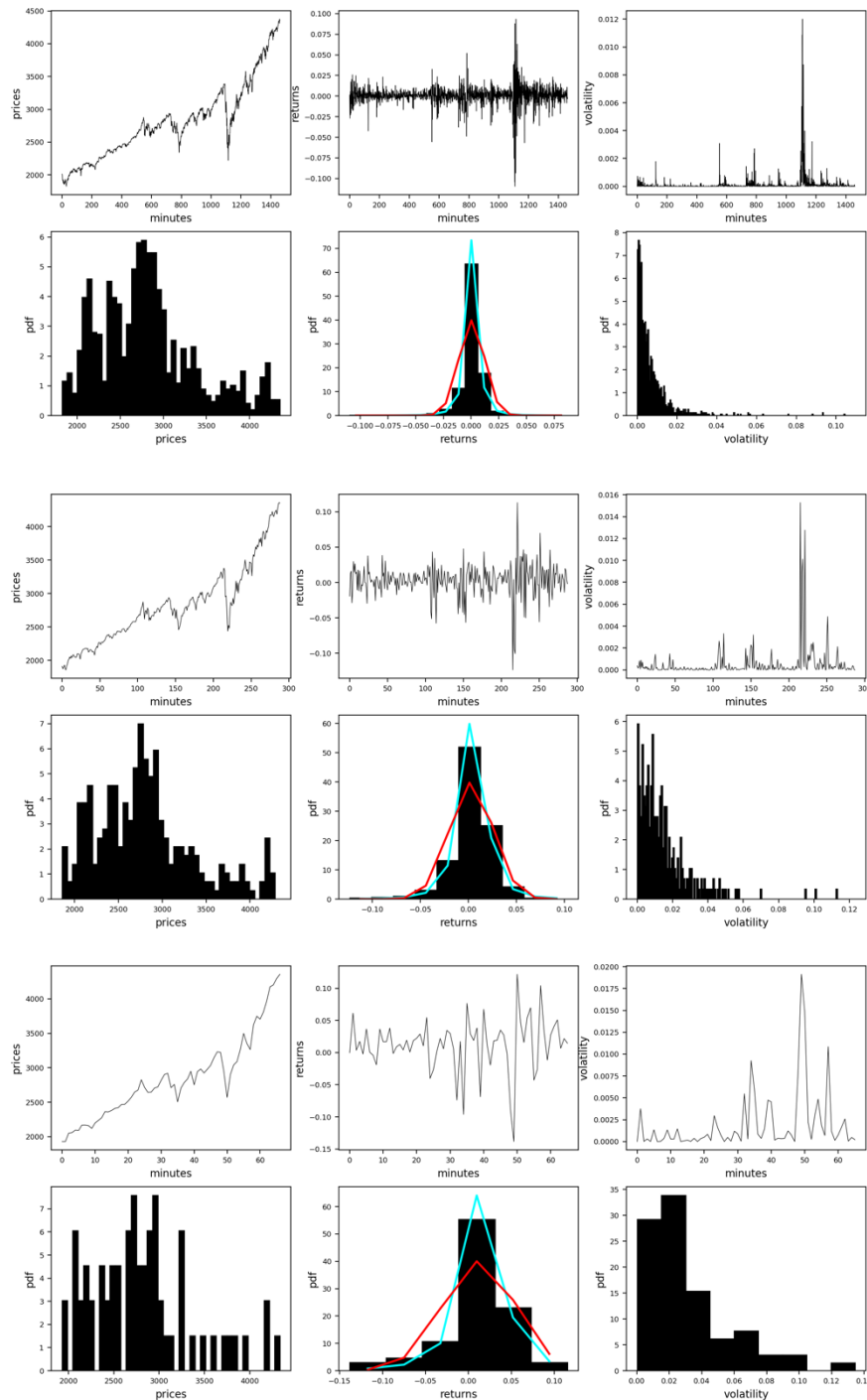


Figure E. 1 Séries temporelles Standard&Poor500 du 01 janvier 2016 au 16 juillet 2021. Du haut vers le bas : période = « journalière », « hebdomadaire », « mensuelle ». La courbe de couleur cyan est l'estimation des données observées par la distribution t-Student. La courbe de couleur rouge correspond l'estimation des données observées par la distribution normale.

Annexe F

Résultats statistiques de l'estimation des paramètres de chaque modèle étudié

Daily data	2016	2017	2018	2019	2020	2021
IGa(x,α,β)	α:1.95, β:3.987e-05	α:1.93, β:1.018e-05	α:1.973, β:4.23e-05	α:1.955, β:7.33e-05	α:0.096, β:2.25e-07	α:1.91, β:7.074e-05
LN(x,μ,σ)	μ:-1.30e-08 σ: 2.343	μ:-2.297e-09 σ: 2.486	μ:-6.468e-09 σ: 2.509	μ:-2.505e-08 σ: 2.33	μ:-529e-09 σ: 2.65	μ:-6.23e-08 σ: 2.16
Ga(x,α,β)	α:0.17, β:0.8085	α:0.11, β:0.7506	α:0.067, β: 0.815	α:0.166, β:0.815	α:0.21, β:197.61	α:0.088, β:0.64

Weekly data	2016	2017	2018	2019	2020	2021
IGa(x,α,β)	α:1.88, β:0.00025	α:2.01, β:8.56e-05	α:1.90, β:0.00026	α:1.36, β:0.00026	α:0.51, β:4.023e-05	α:1.095, β:0.00017
LN(x,μ,σ)	μ:-3.253e-07 σ: 2.187	μ:-3.695e-05 σ: 0.678	μ:-2.29e-07 σ: 2.18	μ:-1.069e-06 σ: 2.015	μ: 1.744e-07 σ:2.471	μ:-2.95e-06 σ: 1.647
Ga(x,α,β)	α:0.53, β:2689.43	α:0.76, β:1/6.95e-05	α:0.42, β:805.03	α:0.48, β:1/ 0.00084	α:0.258, β:87.32	α:0.495, β:1/0.00087

Monthly data	2016	2017	2018	2019	2020	2021
IGa(x,α,β)	α:2.54, β:0.00096	α:0.69, β:6.71e-05	α:1.60, β:0.0015	α:1.55, β:0.00101	α:1.097, β:0.0023	α:1.76, β:0.0011
LN(x,μ,σ)	μ:6.71e-08 σ: 8.14	μ:6.562e-07 σ: 6.252	μ:2.29e-06 σ: 6.873	μ: - 0.00075 σ: 0.618	μ:9.16e-05 σ: 1.77	μ:-0.00033 σ: 0.709
Ga(x,α,β)	α:0.34, β:1/0.00093	α:0.57, β:1/0.00096	α:0.64, β:1/0.0026	α:0.26, β:1/0.0012	α:0.668, β:1/0.0043	α:0.32, β:1/0.00046

Tableau F. 1 Paramètres estimés par maximisation de la vraisemblance des différents modèles. Tableau du haut : estimation des paramètres d'ajustement sur l'échelle temporelle « journalière ». Tableau du milieu : estimation des paramètres d'ajustement sur l'échelle temporelle « hebdomadaire ». Tableau inférieur : estimation des paramètres d'ajustement sur l'échelle temporelle « mensuelle ».

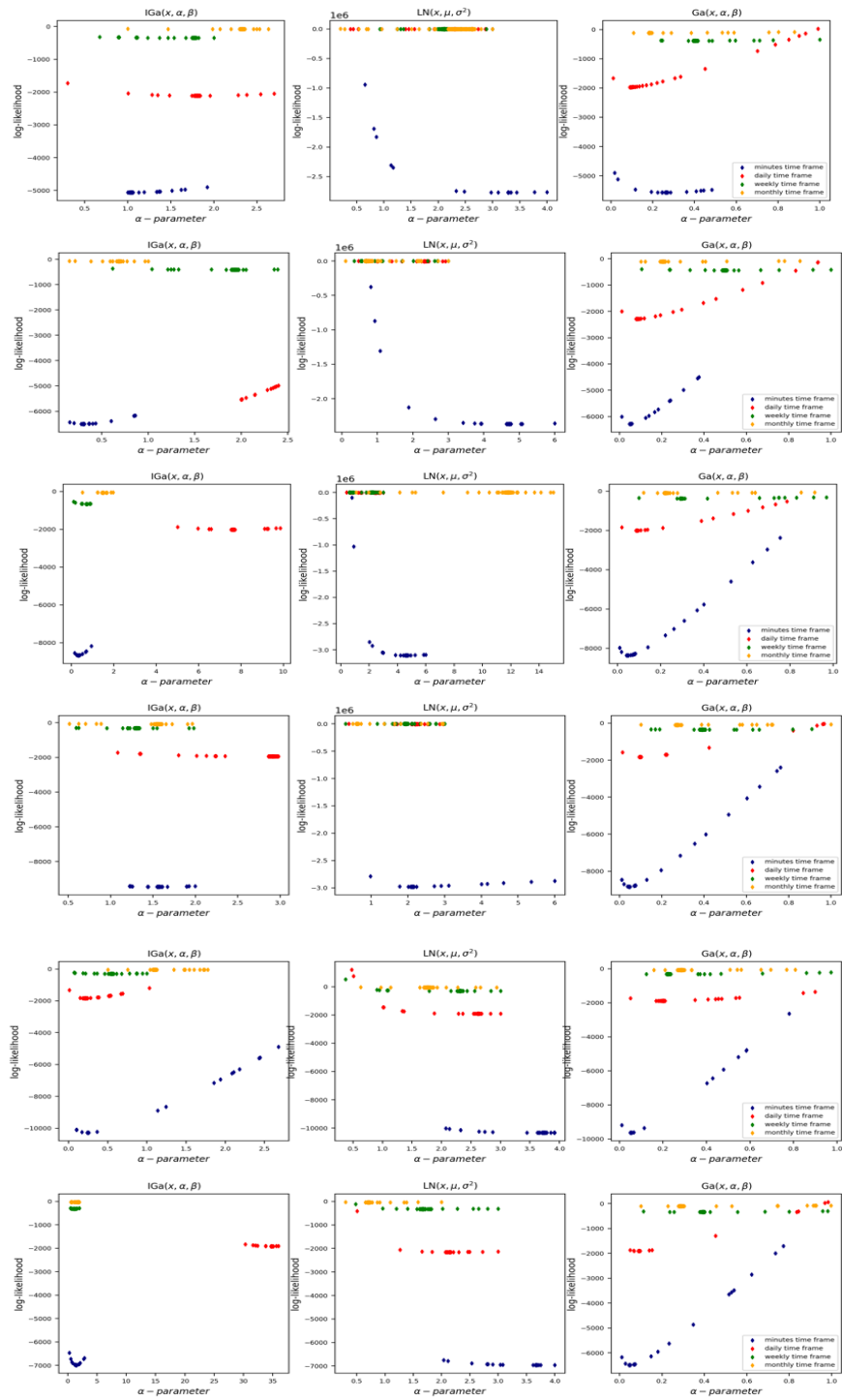


Figure F. 1 Chaque ligne correspond au log-vraisemblance concernant une année respective. Graphique du haut vers le bas, l'échelle temporelle = 2016, 2017, 2018, 2019, 2020, 2021. Graphique de gauche : Log-vraisemblance pour l'estimation des paramètres du modèle IGa. Graphique du milieu : Log-vraisemblance pour l'estimation des paramètres du modèle LN. Graphique de droite : Log-vraisemblance pour l'estimation des paramètres du modèle Ga

Year 2017	minutes	daily	weekly	monthly
Best fit distributions	LN(x,μ,σ)	LN(x,μ,σ)	Ga(x,α,β)	Ga(x,α,β)
SSE from LN	2.43e+09	2.42e+09	4.51e+09	8.28e+07
SSE from IGa	2.55e+09	3.11e+10	3.93e+09	7.66e+07
SSE from Ga	2.55e+09	2.37e+10	2.47e+09	7.60e+07
p-value	0.476	0.489	0.552	0.687

Year 2018	minutes	daily	weekly	monthly
Best fit distributions	LN(x,μ,σ)	IGa(x,α β)	LN(x,μ,σ)	Ga(x,α,β)
SSE from LN	5.00e+08	1.19e+08	1.43e+07	9.01e+06
SSE from IGa	5.51e+08	9.82e+07	1.00e+08	1.05e+07
SSE from Ga	5.54e+08	3.57e+08	2.48e+07	8.74e+06
p-value	0.10	0.359	0.499	0.703

Year 2019	minutes	daily	weekly	monthly
Best fit distributions	LN(x,μ,σ)	LN(x,μ,σ)	LN(x,μ,σ)	IGa(x,α β)
SSE from LN	1.8326e+09	3.50e+07	2.23e+07	4.69e+07
SSE from IGa	1.83296e+09	7.44e+07	3.87e+07	4.45e+07
SSE from Ga	1.83293e+09	3.08e+08	2.26e+07	5.57e+07
p-value	0.058	0.507	0.507	0.738

Year 2020	minutes	daily	weekly	monthly
Best fit distributions	LN(x,μ,σ)	LN(x,μ,σ)	IGa(x,α β)	Ga(x,α,β)
SSE from LN	9.50e+08	1.02e+07	7.01e+05	1.85e+06
SSE from IGa	9.77e+08	2.23e+07	2.96e+05	2.06e+06
SSE from Ga	9.76e+08	1.76e+07	5.58e+06	1.81e+06
p-value	0.149	0.555	0.523	0.647

Year 2021	minutes	daily	weekly	monthly
Best fit distributions	LN(x,μ,σ)	LN(x,μ,σ)	IGa(x,α β)	Ga(x,α,β)
SSE from LN	4.79e+11	3.64e+07	6.93e+06	2.39e+08
SSE from IGa	5.36e+11	7.29e+08	6.42e+06	2.37e+08
SSE from Ga	5.35e+11	7.49e+08	6.58e+06	2.35e+08
p-value	0.132	0.496	0.522	0.443

Tableau F. 2 Comparaison des différents modèles avec la méthode de la somme des erreurs au carré pour les années 2017 à 2021 et sur différentes périodes. Le modèle de distribution qui correspond aux données simulées avec le SSE le plus faible est préféré.

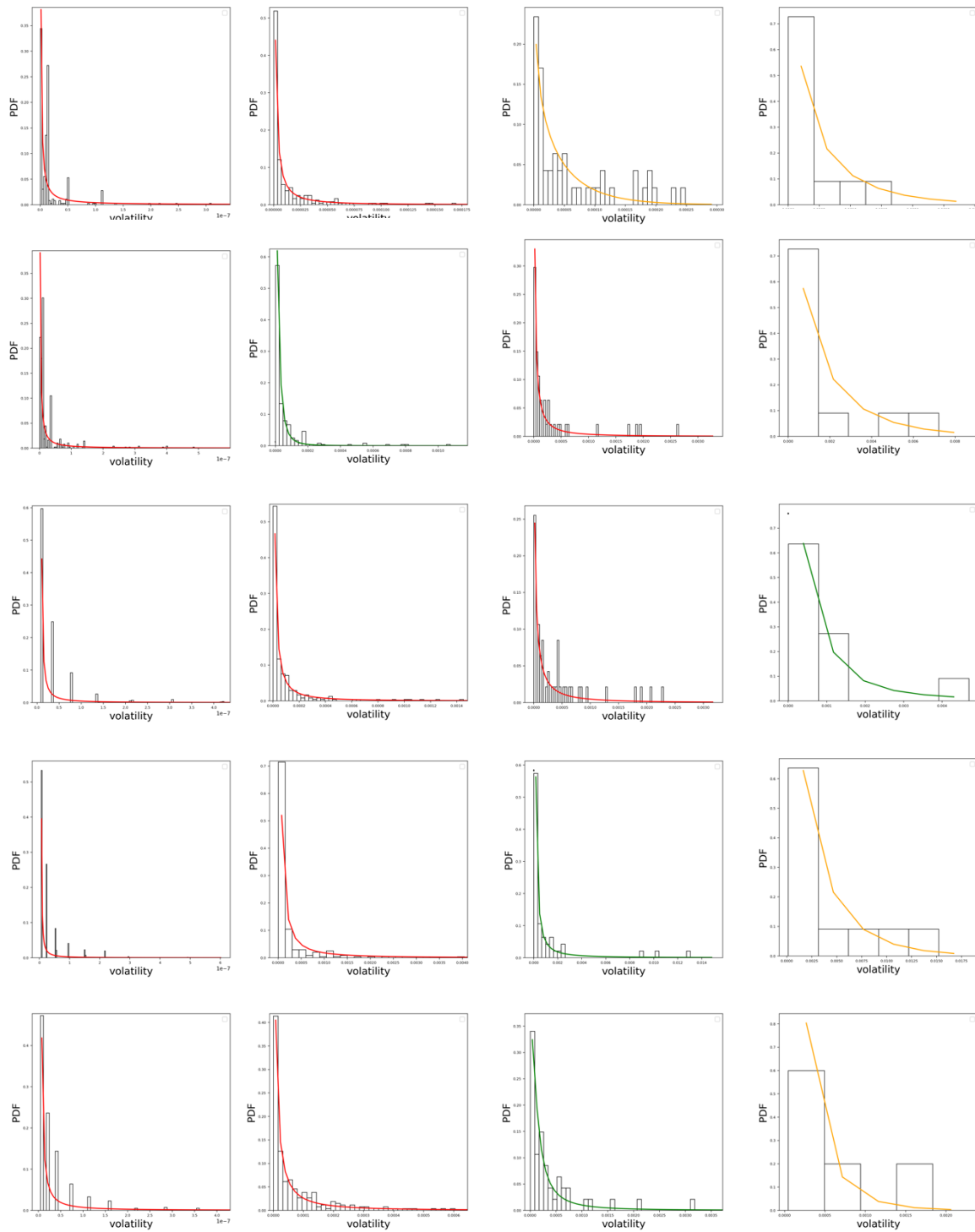


Figure F. 2 Comparaison de la distribution des données empiriques avec la distribution de données simulées issus du modèle le plus significatif. La courbe rouge correspond à la distribution LN, la courbe verte correspond à la distribution IGa et la courbe jaune correspond à la distribution Ga. De gauche à droite, l'échelle de temps = « minute », « journalière », « hebdomadaire » et « mensuel ». Du haut vers le bas, l'année commence en 2017 et se termine en 2021. On observe une transition des statistiques de hautes à basses fréquences, de LN à Ga.

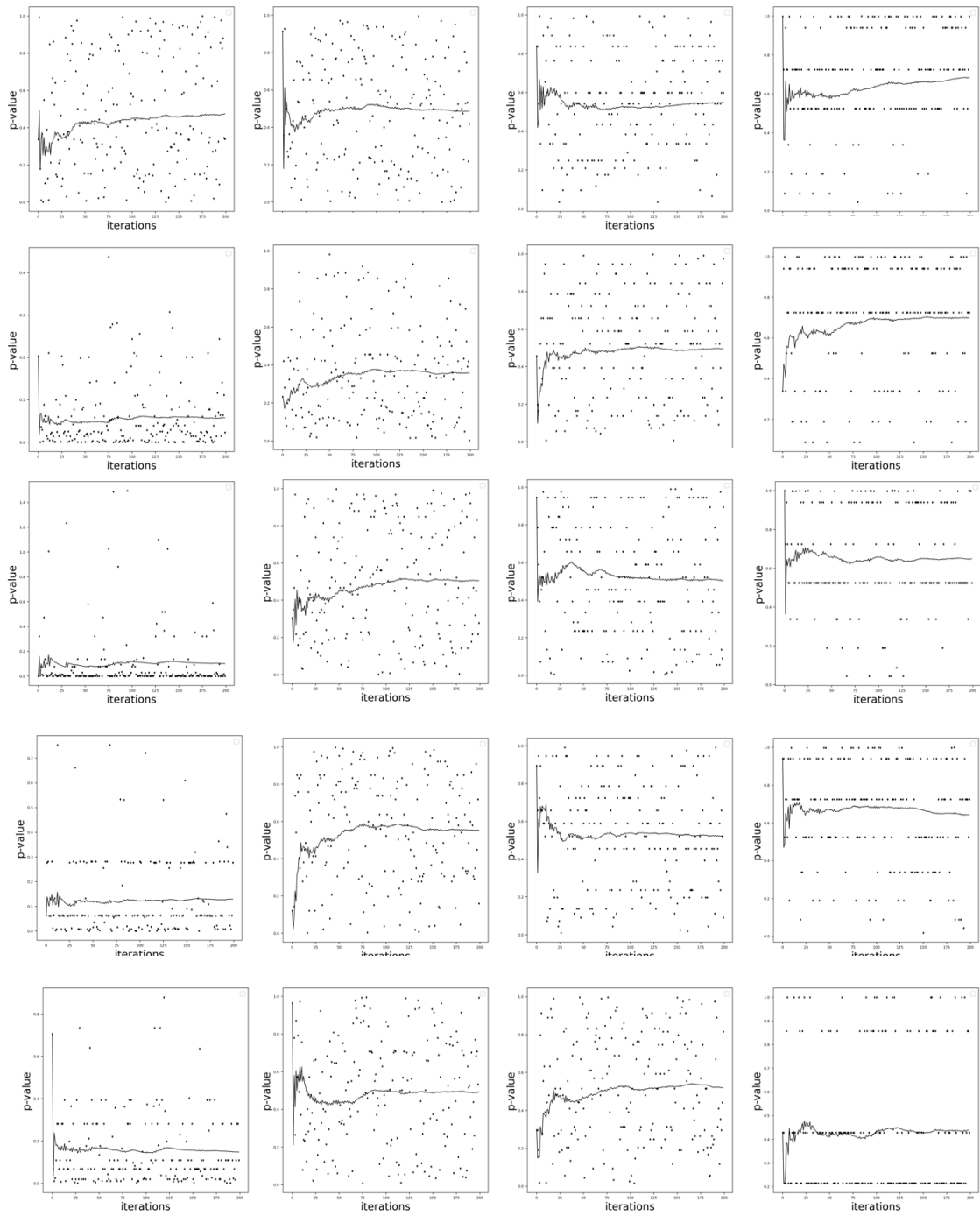


Figure F. 3 Bootstrapping paramétrique de la p-value issu du test de significativité Kolmogorov-Smirnov parmi les modèles d'étude IGa, LN et Ga. De gauche à droite, l'échelle de temps = « minute », « journalière », « hebdomadaire », « mensuel ». De haut en bas, l'année va de 2017 à 2021.

Références

- [1] Abul-Magd, A.Y, Superstatistics in random matrix theory. *Physica A* 361, 41–54, 2006
- [2] Abul-Magd, A.Y, Akemann G., Vivo P. Superstatistical generalizations of Wishart–Laguerre ensembles of random matrices. *J. Phys. A Math. Theor.* 42, 175207, 2009
- [3] Abul-Magd, A.Y, Dietz B., Friedrich, T. & Richer, A. Spectral fluctuations of billiards with mixed dynamics: from time series to superstatistics. *Phys. Rev. E* 77, 046202, 2008
- [4] Amaral L. A. N, Goldberger A. L., Ivanov P.Ch, and Stanley H. E. "Scale-Independent Measures and Pathologic Cardiac Dynamics", *Physical Review Letters*, 81, pp. 2388- 2391, 1998.
- [5] Anderson P.W. More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science*, 177, 1972.
- [6] Ausloos M. and Ivanova K. Dynamical model and nonextensive statistical mechanics of a market index on large time windows. *Phys. Rev. E* 68, 046122, 2003
- [7] Akselrod S. et al. Power Spectrum Analysis of Heart-rate Fluctuation - A Quantitative Probe of Beat-to-beat Cardiovascular Control. *Science* 213, 220–222 (1981).
- [8] Atmanspacher H., Scheingraber H., and Wiedenmann G., “Determination of $f(\alpha)$ for a limited random point set,” *Phys. Rev. A* 40, 3954 (1989).
- [9] Bachelier L. Théorie de la spéculation. *Annales Scientifiques de l’École Normale Supérieure*, III-17:21–86, 1900. [10] Baillie R. T., Bollerslev T. and Mikkelsen H. O. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74:3–30, 1996.
- [11] Bar-yam Yaneer, *Dynamics of Complex Systems (Studies in Nonlinearity) 1st Edition*, ISBN-13: 978-0201557480
- [12] Battiston S., Bonabeau E. and Weisbuch G., Decision making dynamics in corporate boards, arXiv: cond-mat/0209590, 2002.
- [13] Beck C. and Cohen E. G. D. Superstatistics. *Physica A*, 322:267–275, 2003.
- [14] Beck C, Generalized statistical mechanics for superstatistical systems, *Phil. Trans. R. Soc. A* 369, 453–465 doi:10.1098/rsta.2010.0280, 2011

- [15] Beck C., Cohen E. G. D. and Swinney H. L, From time series to superstatistics. *Phys. Rev. E* 72, 056133, 2005
- [16] Beck C., Dynamical foundations of nonextensive statistical mechanics. *Phys. Rev. Lett.* 87, 180601, 2001
- [17] Beck C., Non-extensive statistical mechanics and dparticle spectra in elementary interactions. *Physica A* 286, 164–180, 2000
- [18] Bennett C. H. Logical depth and physical complexity. In R Herken, editor, *The Universal Turing Machine, A Half-Century Survey*, pages 227–257. Oxford University Press, Oxford, 1988.
- [19] Black F. and Scholes M. The valuation of options and Corporate Liabilities. *Journal of Political Economy*, 81:637–654, 1973.
- [20] Briggs K. and Beck C., Modelling train delays with q-exponential functions. *Physica A* 378, 498–504, 2007
- [21] Belankin A, *Fractal Behavior of Complex Systems*, Cientifica Vol.7 Num. 3 pp. 109-128, ENSIME – IPN, ISSN 1665-0654, 2003.
- [22] Bollerslev T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.
- [23] Bouchard J.-P. and Potters M. *Theory of financial risk and derivative pricing*. Cambridge, UK: Cambridge University Press, 2003
- [24] Bouchaud J.P., Potters M. *Financial Applications of Random Matrix Theory: a short review*, *Handbook on Random Matrix Theory*, Oxford University Press.
- [25] Breidt F. J., Crato N., and Lima P. On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83:325–348, 1998.
- [26] Casaleggio A., Bortolan G., Automatic estimation of the correlation dimension for the analysis of electrocardiograms. *Biol Cybern.* 1999
- [27] Chen Leon, and Beck C., A superstatistical model of metastasis and cancer survival. *Physica A* 387, 3162, 2008
- [28] Cheng Q. and Agterberg F.P., “Comparison between two types of multifractal modeling,” *Comput. Geosci.* 28, 1001 (1996).
- [29] Cheng Q., “The gliding box method for multifractal modeling,” *Comput. Geosci.* 25, 1073 (1999).

- [30] Chhabra A.B. and Sreenivasan K.R., “Negative dimensions: Theory, computation, and experiment,” *Phys. Rev. A* 43, 1114 (1991).
- [31] Crutchfield J. P. *The calculi of emergence: computation, dynamics and induction. Physica D: Nonlinear Phenomena*, 75(1-3):11–54, 1994.
- [32] Crutchfield J. P. and Young Karl. Inferring statistical complexity. *Physical Review Letters*, 63:105, 1989.
- [33] Crutchfield J. P. and Shalizi C.R. Thermodynamic depth of causal states: objective complexity via minimal representations. *Physical Review E*, 59(1):275–283, 1999.
- [34] Crutchfield J.P. and Packard N., “Symbolic dynamics of one-dimensional maps: Entropies, finite precision, and noise”, *International Journal of Theoretical Physics*, 21, pp. 433–466, 1982.
- [35] Dale Rick and Spivey M.J., From apples and oranges to symbolic dynamics: a framework for conciliating notions of cognitive representation. *Journal of Experimental & Theoretical Artificial Intelligence*, Vol. 17, No. 4, 317-342, December 2005
- [36] Dantas E. M. et al, Comparison between symbolic and spectral analyses of short-term heart rate variability in a subsample of the ELSA-Brasil study. *Physiol. Meas.* 36, 2119–2134 (2015).
- [37] Dacorogna M.M., Genc R., Muller U. A., Olsen R.B., and Pictet O.V.. *An Introduction to High-Frequency Finance*. CA: Academic Press, San Diego, 2001.
- [38] Davidovits Paul, Chapter 10 – Thermodynamics - *Physics in Biology and Medicine (Fifth Edition)*, 2019
- [39] Ding Z., Granger C., and Engle R. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106, 1993.
- [40] Ebeling W., Rateitschak K, Symbolic dynamics, Entropy and Complexity of the Feigebaum Map at the Accumulation Point, *Discrete Dynamics in Nature and Society*, Vol.2, pp. 187-194.
- [41] Eckmann J.P., Kamphorst S.O., Ruelle D., Ciliberto S., Lyapunov exponents from time series. *Physical Review A* 34, 4971-4979, 1986
- [42] Eckmann J.P., Ruelle D. Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics* 57, 617–650, 1985

- [43] Einstein A. On the Movement of Small Particles Suspended in a Stationary Liquid Demanded by the Molecular-Kinetic Theory of Heat. *Ann. Physik*, 17:549–560, 1905
- [44] Engle R. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1007, 1982.
- [45] Fernandez-Rodriguez F., Sosvilla-Rivero S., Andrada-Félix, J., 2005. Testing chaotic dynamics via Lyapunov exponents. *Journal of Applied Econometrics* 20, 901-930.
- [46] Foote R. Mathematics and complex systems. *Science*, 318:410–412, October 2007.
- [47] Frisch U. and Parisi G., “Fully developed turbulence and intermittency,” in *Turbulence and Predictability of Geo- physical Flows and Climate Dynamics*, edited by M. Ghil, R. Benzi, and G. Parisi pp. 84–88 (North-Holland, Amsterdam, 1985)
- [48] Gautrais J., Jost C. & Theraulaz Guy, Key behavioral factors in a self-organized fish school odell, *Ann. Zool. Fennici* 45: 415-428, ISSN 0003-455X, Helsinki 29 October 2008.
- [49] Gell-Mann M. and Lloyd S. Information measures, effective complexity, and total information. *Complexity*, 2(1):44–52, 1996.
- [50] Gell-Mann M. What is complexity. *Complexity*, 1:1, 1995.
- [51] George M. W. and Ismagilov R. F. Complexity in chemistry. *Science*, 284:89–92, April 1999.
- [52] Ghashghaie S., Breymann W., Peinke J., Talkner P., and Dodge Y. Turbulent Cascades in Foreign Exchange Markets. *Nature*, 381:767–770, 1996.
- [53] Gibbs J.W., *Elementary principles in statistical mechanics*, Yale University, The rational foundation of thermodynamics, 1902
- [54] Goldenfield N. and Kadanoff L.P. Simple lessons from Complexity, 284:87-89. April 1999. Editorial. No man is an island. *Nature Physics*, 5:1, 2009.
- [55] Govindan R. B., Bunde A. and Havlin S., Volatility in atmospheric temperature variability, arXiv: cond-llat/0209687.
- [56] Grassberger P. and Procaccia I. "Measuring the Strangeness of Strange Attractors". *Physica D: Nonlinear Phenomena*. 9 (1–2): 189–208, 1983
- [57] Grassberger P. Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics*, 25(9):907–938, 1986.

- [58] Grassberger P., "Generalized dimensions of strange attractors," *Phys. Lett. A* 97, 227 (1983).
- [59] Guzzetti S. et al., *Symbolic Dynamics of Heart Rate Variability: A Probe to Investigate Cardiac Autonomic Modulation*. *Circulation* 112, 465–470 (2005).
- [60] Hadamard J. *Les surfaces à courbures opposées et leurs lignes géodésiques*, *Journal de Mathématiques*, 1898.
- [61] Halsey T. C., Jensen M. H., Kadanoff L. P., Procaccia I., and Shraiman B. I., "Fractal measures and their singularities: the characterization of strange sets," *Phys. Rev. A* 33, 1141 (1986).
- [62] Hentschel H.G.E., Procaccia I, *The Infinite Number of Generalized Dimensions of Fractals and Strange Attractors*, *Physica* 8D, 435-444, 1983
- [63] Herbert A. S. *The architecture of complexity*. *Proceedings of the American Philosophical Society*, 106(6):467–482, 1962
- [64] Heston S. L. *A Closed Solution For Options With Stochastic Volatility, With Application to Bond and Currency Options*. *Review of Financial Studies*, 6(2):327–343, 1993.
- [65] Hunsaker N. C., *Hadamard's Theory of Geodesics on Surfaces of Negative Curvature*, A minor thesis presented to the Faculty of the Rice Institute, 1941
- [66] Ivanova V. S., Balankin A. S., Bunin A., & Oksogoev A., *Synergetics and Fractals in Material Science*. Nauka, Moscow, 1994.
- [67] Kennel M.B. and Buhl M., "Estimating good discrete partitions from observed data: symbolic false nearest neighbors", *Physical Review Letters*, 91, 084102, 2003.
- [68] Klein M. J., *The Physics of J. Willard Gibbs in his time*, *American Institute of Physics*, 43, 9, 40 (1990)
- [69] Klonowski W. *Signal and Image Analysis Using Chaos Theory and Fractal Geometry*, Polish Academy of Sciences, 2000.
- [70] Kozlov V.I. and Markov V.v., *Sun Cycles as Interchange of Low dimension Attractors in Chaotic Dynamics of Solar Activity*, Institute of Cosmo physical Research and Aeronoll1Y, 31 Lenin Ave., 677891 Yakutsk, Russia, 2003.
- [71] Kurths J., Voss A., Sapanin P., Witt A., Kleiner H. and Wessel N., "Quantitative analysis of heart rate variability", *Chaos*, 5, pp. 88–94, 1995.

- [72] Lanford O.E., Entropy and equilibrium states in classical statistical mechanics, *Statistical Mechanics and Mathematics Problems* (1973).
- [73] Laubrich T., Ghasemi F., Peinke J. and Kantz H., Statistical analysis of wind speed fluctuation and increments of non-stationary atmospheric boundary layer turbulence, 2008
- [74] Lloyd S. and Pagels H. Complexity as thermodynamic depth. *Annals of Physics*, 188:186–213, November 1988.
- [75] Lloyd S. Measures of complexity: a non-exhaustive list. *Control Systems Magazine, IEEE*, 21:7–8, 2001.
- [76] Ma Tao, Serota R.A., A model for stock returns and volatility, *Physica A* 398 89–115, 2014
- [77] MacKay R.S. *Nonlinearity in complexity science*. 2008.
- [78] Mandelbrot B. B. Long-run linearity, locally Gaussian processes, H-spectra and infinite variance. *International Economic Review*, 10:82–111, 1969.
- [79] Mandelbrot B. B. *Fractals and chaos: the Mandelbrot set and beyond*, ISBN 978-1-4757-4017-2.
- [80] Mantegna R. N. and Stanley H. E. Scaling Behavior in the Dynamics of an Economic Index. *Nature*, 376:46–49, 1995.
- [81] Mantegna R. N. and Stanley H. E.. *Stock Market Dynamics and Turbulence: Parallel Analysis of Fluctuation Phenomena*. *Physica A*, 239:255–266, 1997.
- [82] Maxwell J.C., on Governors, M.A., F.R.S.S.L. & E. March 5, 1868
- [83] Maxwell J.C., *Illustrations of the Dynamical Theory of Gases* Originally published in *Phil. Mag.*, Vol. 19, pp. 19–32; Vol. 20, pp. 21–37 ; reprinted in *The Scientific Papers of James Clerk Maxwell* (ed. W. D. NIVEN), Cambridge University Press, 1890 Vol. I, pp. 377–409, 1860
- [84] Mills T. Stylized facts of the temporal and distributional properties of daily FTSE returns. *Applied Financial Economics*, 7:599–604, 1997.
- [85] Morin Edgar, *La méthode, La Nature de la nature*, t. 1, Paris, Le Seuil, coll. « Points », 414 p. (ISBN 978-2-02-005771-4), 1977
- [86] Nilson D. and Johansson A., Social influence during the initial phase of a fire evacuation—Analysis of evacuation experiments in a cinema theatre, *Fire Safety Journal* 44(1):71-79 DOI: 10.1016/j.firesaf.2008.03.008

- [87] Ohtaki Y. and Hasegawa H. H., Superstatistics in econophysics, 2003
- [88] Osborne M. F. M., Brownian motion in the stock market, US Naval Research Laboratory, Washington 25, DC
- [89] Pareto V. Cours d'Economie Politique. Lausanne and Paris, 1897.
- [90] Perseguini N. M. et al., Spectral and symbolic analysis of the effect of gender and postural change on cardiac autonomic modulation in healthy elderly subjects. *Braz. J. Med. Biol. Res.* 44, 29–37 (2011).
- [91] Paladin G. and Vulpiani A., "Anomalous scaling laws in multi- fractal objects," *Phys. Rep.* 156, 147 (1987).
- [92] Pavlov Dimitrii, Kasumyan A.O., Patterns and mechanisms of schooling behavior in fish: A review, *Journal of Ichthyology*: S163-S231
- [93] Porporato A., Vico G., and Fay P. A., Superstatistics of hydro-climatic fluctuations and interannual ecosystem productivity. *Geophys. Res. Lett.* 33, L15402, 2006
- [94] Porta A. et al. Assessment of cardiac autonomic modulation during graded head-up tilt by symbolic analysis of heart rate variability. *Am J Physiol Heart Circ Physiol* 293, H702–H708 (2007).
- [95] Rabassa P. and Beck C. Superstatistical analysis of sea level fluctuations. *Physica A*, 417:18, 2015.
- [96] Renyi A., *Probability Theory* (North-Holland, Amsterdam, 1970).
- [97] Reynolds A. Superstatistical mechanics of tracer-particle motions in turbulence. *Phys. Rev. Lett.* 91, 084503, 2003
- [98] Rind D. Complexity and climate. *Science*, 284:105–107, April 1999.
- [99] Rizzo S. and Rapisarda A., Environmental atmospheric turbulence at Florence airport. *AIP Conf. Proc.* 742, 176–181, 2004
- [100] Salat Hadrien, Murcio Roberto, and Elsa Accurate Multifractal methodology, 10.1016/j.physa.2017.01.041, 2016
- [101] Shannon Claude, *A Mathematical Theory of Communication*, The Bell System Technical Journal, Vol. 27, pp. 379-423, 623-656, July, October 1948.
- [102] Schlaifer and Raiffa, *Applied statistical decision theory*, Division of Research, Graduate School of Business Administration, Harvard University, ISBN: 0875840175, 9780875840178, 1961

- [103] Shintani M., Linton O., Nonparametric neural network estimation of Lyapunov exponents and a direct test for. *Journal of Econometrics* 120, 1-33, 2004
- [104] Smale Steve, Finding a Horseshoe on the beaches of Rio, *The Mathematical Intelligencer*, DOI: 10.1007/BF03024399
- [105] Smale Steve, The Story of the Higher Dimensional Poincaré Conjecture (What Actually Happened on the Beaches of Rio), *From topology to Computation: Proceeding of the Smalefest* pp 27-40.
- [106] Smith Jr. T.G., Marks W.B., Lange G.D., Sheriff Jr. W.H., and Neale E.A. A fractal analysis of cell images. *J.Neurosci.Meth.* 27, 173-180. 1990
- [107] So P., Barreto Ernest, and Hunt B. R. "Box-counting dimension without boxes: Computing D from average expansion rates", *Physical Review*, E60, pp. 378-385, 1999.
- [108] Soulier P., Kulik R., Limit theorems for long memory stochastic volatility models with infinite variance: partial sums and sample covariance,
- [109] Straeten E.V.d and Beck C, Superstatistical distributions from a maximum entropy principal, *Physical review E* 78, 051101 (2008)
- [110] Sub B., Alencar A. M., Frey U., Ivanov P. Ch., Buldyrev S. V., Majumdar A., Stanley H. E., Dawson Ch. A., Krenz G. S. and Mishima M. "Fluctuations, Noise and Scaling in the Cardio- Pulmonary System", *Fluctuations and Noise Letters*, 1, 2000.
- [111] Sumiyoshi Abe, Beck Christian, and Ezechiel G. D. Cohen, Superstatistics, thermodynamics and fluctuations, wrking paper, Department of Physical Engineering,
- [112] Sumiyoshi Abe, Beck Christian, and Ezechiel G. D. Cohen, Fokker-Planck theory of superstatistics.
- [113] Takens F. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence*, Rand D, Young L. (Eds.). Springer-Verlag, Berlin, 366-381, 1981
- [114] Taylor S. J. *Modelling Financial Time Series*. Wiley, 1986.
- [115] Tobaldini E. et al., Symbolic analysis detects alterations of cardiac autonomic modulation in congestive heart failure rats. *Auton. Neurosci. Basic Clin.* 150, 21–26 (2009).
- [116] Tsallis C., Mendes R., Plastino A. The role of constraints within generalized non extensive statistics. *Physica* 261A 534–554, 1998

[117] Wasserman Larry and Liu Han, Statistical Machine Learning, Chapter 12 – Bayesian Inference, p302, p.303, p.311, p.312, p.315.

[118] Weibull J.W., Logfren Karl-Gustaf, PerssonTorsten, Markets with Asymmetric Information: The Contributions of George Akerlof, Michael Spence and Joseph Stiglitz, Scand. J. of Economics 104(2), 195-211, 2002

[119] Werner B. T. Complexity in natural landform patterns. Science, 284:102–104, April 1999.

[120] Wilk G. and Wlodarczyk Z. Interpretation of the non-extensive parameter q in some applications of Tsallis statistics and Levy distributions. Phys. Rev. Lett. 84, 2770-2773doi:10.1103/PhysRevLett.84.2770, 2000

[121] Wu Yihong, Vurdu Sergio, Rényi Information Dimension: Fundamental Limits of Almost Lossless Analog Compression, IEEE transactions theory, Vol. 56, No. 8, August 2010.

[122] Xu Dan, Superstatistics and symbolic dynamics of share price returns on different time scales, School of Mathematical Sciences, Queen Mary University of London, January 18, (p106, 109, 112, 113), 2017

[123] Zmeskal O., Nezadal M. Buchnicek M. Fractal–Cantorian geometry, Hausdorff dimension and the fundamental laws of physics Chaos, Solitons & Fractals, 17, pp. 113-119.

Table des Figures

Figure 1. 1 Exemples de déplacements collectifs résultants de processus d’auto-organisation du fait d’interactions et d’interdépendance entre agents d’un système. (a) Formation d’un banc de poissons cherchant à fuir un prédateur. (b) Vol groupé d’Étourneau. (c) Croisement de piétons à Shibuya, Tokyo. (d) Construction d’un pont de fourmis servant de raccourcis aux chasseuses. (<i>istockphoto.com</i>)	20
Figure 1. 2 Exemple de forme régulières symétriques et ordonnées. (1) Tronc d’arbre. (b) Escargot de mer. (c) Nid d’abeilles. (d) Zèbre. (<i>istockphoto.com</i>)	21
Figure 1. 3 Modèle de panique dans un espace clos et répartition des individus en zone de stress [11].	23
Figure 1. 4	27
Figure 1. 5 Self-similaire fractales. Similaire apparence sur des échelles plus petites. La structure de droite est similaire à l’ensemble de la structure représentée à gauche.	29
Figure 1. 6 Construction itérative de courbes fractales dites self-affine dont leur dimension respective dépend du facteur d’échelle d’observations. (https://www.rgbstock.com/photo/omq&JIC/Fractal+Tree+8)	29
Figure 1. 7 Figure provenant de “The Computational Beauty of Nature: Computer Exploration of Fractals, Chaos, Complex Systems and Adaptation”. Copyright 1998-2000 by Gary William Flake. .	31
Figure 2. 1 Exemple d’une mise à l’échelle d’un système par recouvrement de ce dernier avec des formes géométriques de faible diamètre ou de faible longueur (cercle, carré) pour déterminer sa dimension avec la méthode de Hausdorff. (<i>google.com/images.com</i>)	44
Figure 2. 2 Exemple d’une mise à l’échelle d’un système avec un système de grille pour déterminer la dimension avec la méthode de « box-counting ». (<i>google.com/images.com</i>)	45
Figure 2. 3 Prix et rendements logarithmiques historiques du Standard&Poor500 entre janvier 2016 et mars 2021 sur l’échelle de temps minute.	68
Figure 2. 4 Prix et rendements logarithmiques historiques du Standard&Poor500 entre janvier 2016 et mars 2021 sur l’échelle de temps 4-heures.....	68
Figure 2. 5 Prix et rendements logarithmiques historiques du Standard&Poor500 entre janvier 2016 et mars 2021 sur l’échelle de temps journalière.	68
Figure 2. 6 Probabilités des rendements logarithmiques de la série chronologique sur l’échelle minute et pour les partitions de longueur $L = [2,4,6,8]$. Par conséquent, chaque série a respectivement 4, 16, 64 et 256 types de symboles possibles. En bleu, nous représentons les séquences de symboles dynamiques du Standard & Poor500, en rouge nous représentons une séquence de symboles dynamiques aléatoires.	73

Figure 2. 7 Probabilités des rendements logarithmiques de la série chronologique sur l'échelle 4-hours et pour les partitions de longueur $L = [2,4,6,8]$. Par conséquent, chaque série a respectivement 4, 16, 64 et 256 types de symboles possibles. En bleu, nous représentons les séquences de symboles dynamiques du Standard & Poor500, en rouge nous représentons une séquence de symboles dynamiques aléatoires. 74

Figure 2. 8 Probabilités des rendements logarithmiques de la série chronologique sur l'échelle journalière et pour les partitions de longueur $L = [2,4,6,8]$. Par conséquent, chaque série a respectivement 4, 16, 64 et 256 types de symboles possibles. En bleu, nous représentons les séquences de symboles dynamiques du Standard & Poor500, en rouge nous représentons une séquence de symboles dynamiques aléatoires. 75

Figure 2. 9 En bleu, graphique Box-Plot représentant les caractéristiques statistiques des séquences de symboles de longueur $L = 2$ sur différentes échelles de temps (minutes, 4 heures, par jour) de la dynamique des retours logarithmiques de Standard & Poor500. En gris, graphique Box-Plot représentant les caractéristiques statistiques de séquences de symboles de longueur $L = 2$ sur différentes échelles de temps (minutes, 4 heures, par jour) de symboles dynamiques parfaitement aléatoires. 81

Figure 2. 10 En bleu, graphique Box-Plot représentant les caractéristiques statistiques des séquences de symboles de longueur $L = 4$ sur différentes échelles de temps (minutes, 4 heures, par jour) de la dynamique des retours logarithmiques de Standard & Poor500. En gris, graphique Box-Plot représentant les caractéristiques statistiques de séquences de symboles de longueur $L = 4$ sur différentes échelles de temps (minutes, 4 heures, par jour) de symboles dynamiques parfaitement aléatoires. 82

Figure 2. 11 En bleu, graphique Box-Plot représentant les caractéristiques statistiques des séquences de symboles de longueur $L = 6$ sur différentes échelles de temps (minutes, 4 heures, par jour) de la dynamique des retours logarithmiques de Standard & Poor500. En gris, graphique Box-Plot représentant les caractéristiques statistiques de séquences de symboles de longueur $L = 6$ sur différentes échelles de temps (minutes, 4 heures, par jour) de symboles dynamiques parfaitement aléatoires. 82

Figure 2. 12 En bleu, graphique Box-Plot représentant les caractéristiques statistiques des séquences de symboles de longueur $L = 8$ sur différentes échelles de temps (minutes, 4 heures, par jour) de la dynamique des retours logarithmiques de Standard & Poor500. En gris, graphique Box-Plot représentant les caractéristiques statistiques de séquences de symboles de longueur $L = 8$ sur différentes échelles de temps (minutes, 4 heures, par jour) de symboles dynamiques parfaitement aléatoires. 83

Figure 2. 13 Histogramme du ratio \emptyset selon les échelles de temps (minute, 4-hours, journalière) et selon l'échelle spatiale des séquences de symboles ($L = \{2,4,6,8\}$). 84

Figure 2. 14 Entropie de Rényi de la dynamique des séquences de symboles Standard&Poor500 pour $L = \{2,4,6,8\}$ selon l'échelle de temps minute. 88

Figure 2. 15 Entropie de Rényi de la dynamique des séquences de symboles Standard&Poor500 pour $L = \{2,4,6,8\}$ selon l'échelle de temps 4-heures. 88

Figure 2. 16 Entropie de Rényi de la dynamique des séquences de symboles Standard&Poor500 pour $L = \{2,4,6,8\}$ selon l'échelle de temps journalière.	89
Figure 2. 17 Comparaison des entropies de Rényi de la dynamique des séquences de symboles Standard & Poor500 pour $L = 2$ (a) et pour $L = 8$ (b) selon différentes échelles de temps.	90
Figure 2. 18 Comparaison des dimensions de Rényi de la dynamique des séquences de symboles Standard & Poor500 pour $L = 2$ (a) et pour $L = 8$ (b) selon différentes échelles de temps.	90
Figure 2. 19 (a) Comparaison des courbes αq versus q définies par l'équation (3.19) selon différentes échelles de temps ; (b) comparaison des fonctions $f(\alpha(q))$ contre αq pour les valeurs d'échelle spatiale allant de 2 à 8 et q allant de -30 à 30 selon différentes échelles de temps.	91
Figure 3. 1 Distribution de Maxwell-Boltzmann pour un gaz idéal selon différentes températures données (https://commons.wikimedia.org/wiki/File:Maxwell-Boltzmann-Distribution-simple-axis-labels.svg).	99
Figure 3. 2 Représentation schématique d'un système comme ensemble physique constitué d'un ensemble de sous-systèmes.	100
Figure 3. 3 Représentation du comportement de molécules d'un gaz selon l'activité thermique du système. La température est liée à l'énergie cinétique moyenne des composants du système.	101
Figure 3. 4 Représentation d'une particule Brownienne dans un environnement spatio-temporel inhomogène du fait d'une température inverse βi différente selon les régions du système (Figure inspirée de l'article [14]).	107
Figure 3. 5 Expérience sous l'hypothèse d'un système conservatif. Graphique supérieur gauche : moyenne lissée de la probabilité de profit postérieur du macro-état. Graphique supérieur droite : distribution de la probabilité de profit du macro-état. Graphique inférieur gauche : évolution du nombre de combinaisons possibles de micro-états avec une probabilité marginale égale à 0,5. Graphique inférieur droite : évolution de la probabilité postérieure d'opportunité de profit de chaque micro-états. La courbe rouge représente la probabilité moyenne de l'ensemble des micro-états.	117
Figure 3. 6 Cas d'un système conservatif. A gauche : répartition de l'opportunité de profits sous forme d'histogrammes. Modélisation de l'opportunité de profit selon différentes statistiques (bleu : t-étudiants, jaune : Laplace, vert : Cauchy, rouge : Pearson et violet : norme). A droite : même figure mais agrandie.	119
Figure 3. 7 Expérience sous l'hypothèse d'un système dissipatif. Graphique supérieur gauche : moyenne lissée des probabilités postérieures de profit des macro-états locaux (dit « coarse-grained »). Graphique supérieur droite : distribution de la probabilité postérieure des opportunités de profits de chaque macro-état local. Graphique inférieur gauche : évolution du nombre de micro-états avec une probabilité marginale de 0.5 pour chaque macro-état local. Graphique inférieur droite : évolution de la probabilité postérieure de profit moyenne des macro-états locaux.	121

Figure 3. 8 Superposition de la distribution des macro-états local d'un système dissipatif. A gauche : répartition de l'opportunité de profit du système sous forme d'histogrammes. Modélisation de l'opportunité de profit selon différentes statistiques (bleu : t-étudiants, jaune : Laplace, vert : Cauchy, rouge : Pearson et violet : norme). A droite : même figure mais agrandie.	122
Figure 3. 9 Distribution de probabilité des opportunités de profits de chaque macro-état local en histogramme avec la distribution de t-student en bleu. En haut à gauche : macro-état local ω_1 avec 750 micro-états. En haut à droite : macro-état local ω_2 avec 225 micro-états. En bas à gauche : macro-état local ω_3 avec 150 micro-états. En bas à droite : macro-état local ω_1 avec 425 micro-états.	124
Figure 3. 10 Variation absolue des rendements logarithmiques du S&P500 selon différente échelle temporelle. En haut à gauche : « minute ». En haut à droite : « heure ». En bas à gauche : « 4-heures ». En bas à droite : « journalière ».....	130
Figure 3. 11 Graphique log-log de CCDF en fonction de la variance. Courbes noires : tracé log-log. Droite rouge : Régression linéaire du $\log_{10}(\text{CCDF})$. En haut à gauche : période = « minute », $k = -2.071$. En haut à droite : période = « heure », $k = -1,853$. En bas à gauche : période = « 4 heures », $k = -1,683$. En bas à droite : période = « journalière », $k = -1.051$	131
Figure 3. 12 Fonction d'autocorrélation des valeurs de la variance du Standard&Poor500. En haut à gauche : « minutes », En haut à droite : « heures », En bas à gauche : « 4 heures », En bas à droite : « journalière ». La courbe bleue est la zone d'intervalle de confiance de 5%.....	132
Figure 3. 13 Histogramme de la variance du Standard&Poor500. En haut à gauche : séries chronologiques sur l'échelle « minute ». En haut à droite : séries chronologiques sur l'échelle « heure ». En bas à gauche : séries chronologiques sur l'échelle « 4 heures ». En bas à droite : séries chronologiques sur l'échelle « journalière ». Ligne bleue : distribution des données observées avec la distribution inverse gamma. Ligne rose : distribution des données observées avec la distribution log-normal.....	133
Figure 3. 14 Représentations graphiques des distributions de propositions issues des 15 meilleurs paramètres statistiques α, β du modèle $M1 = \text{IGa}$ qui maximisent la probabilité postérieure. Ici, l'échelle de temps d'étude est « minute ». A gauche : probabilité postérieure. Milieu : probabilité marginale. A droite : probabilité a priori.	137
Figure 3. 15 Représentations graphiques des distributions de propositions issues des 15 meilleurs paramètres statistiques α, β du modèle $M2 = \text{LN}$ qui maximisent la probabilité postérieure. Ici, l'échelle de temps d'étude est « minute ». A gauche : probabilité postérieure. Milieu : probabilité marginale. A droite : probabilité a priori.	138
Figure 3. 16 Distributions des facteurs de Bayes (BF) des séries de variables de variance simulées (c.-à-d. σ_2) conformément aux modèles estimés $M1$ et $M2$. En haut à gauche : période = « minute ». En haut à droite : période = « heure ». En bas à gauche : période = « 4 heures ». En bas à droite : période = « journalière ». Les histogrammes bleus correspondent à $\text{BF} > 1$ et les histogrammes roses à $\text{BF} < 1$. La ligne verticale rouge correspond à la médiane des facteurs bayésiens, et la ligne bleue en pointillé est fixée à 1.....	140

Figure 3. 17. Comparaison de la distribution de la densité de probabilité entre les données observées D_n et les log-retours simulés du modèle 1 et du modèle 2. De haut en bas : la période va de « minute » à « journalière ». Histogrammes blancs avec bordure noire : distribution des données empiriques. Histogrammes fuchsia représente la distribution des données simulées issues du Modèle 1. Histogrammes cyan représente la distribution des données simulées issues du Modèle 2.....	142
Figure 3. 18 Séries temporelles Standard&Poor500 sur l'échelle « minute » du 1er janvier 2016 au 16 juillet 2021.	145
Figure 3. 19 Étude statistique pour l'année 2016. Sur le graphique supérieur et inférieur, de gauche à droite, l'échelle temporelle = « minute », « journalière », « hebdomadaire », « mensuelle ».....	147

Liste de Tableaux

Tableau 2. 1 Probabilités conditionnelles des séquences de symboles de longueur $L = \{2,4,6\}$ sur l'échelle de temps minute.....	78
Tableau 2. 2 Probabilités conditionnelles des séquences de symboles de longueur $L = \{2,4,6\}$ sur l'échelle de temps 4-hours.....	78
Tableau 2. 3 Probabilités conditionnelles des séquences de symboles de longueur $L = \{2,4,6\}$ sur l'échelle de temps journalier.....	79
Tableau 2. 4 Comparaison pour différentes longueurs de séquence L et pour différentes échelles de temps le rapport de longueur interquartile (IQR) que nous définit comme suit : $\emptyset = \text{length of IQR from SP500} / \text{length IQR from random series}$	84
Tableau 3. 1 Résultats statistiques des données simulées et de leur distribution (Cas d'un système conservatif).....	118
Tableau 3. 2 Comparaison de différentes modèles statistiques avec la distribution des des opportunités de profits simulés. Méthode de la somme des erreurs au carré es utilisée. Le modèle de distribution qui modélise les données observées avec le SSE le plus faible doit être préféré.	119
Tableau 3. 3 Résultats statistiques des données simulées et de leur distribution (Système dissipatif).	121
Tableau 3. 4 Comparaison de différentes modèles statistiques avec la distribution des opportunités de profits simulés. Méthode de la somme des erreurs au carré est utilisée. Le modèle de distribution qui modélise les données observées avec le SSE le plus faible doit être préféré.	123
Tableau 3. 5 Résultats Statistiques de l'estimation des paramètres de la distribution des macro-états locaux avec la loi t-student.....	124
Tableau 3. 6 Test de stationnarité Kwiatkowski-Phillips-Schmidt-Shin (KPSS) sur les rendements logarithmiques absolus. L'hypothèse nulle : la série temporelle est stationnaire. Hypothèse alternative : la série temporelle a une racine unitaire (non stationnaire). Si la statistique de test est supérieure à la valeur critique, nous rejetons l'hypothèse nulle.....	132
Tableau 3. 7 Estimation des paramètres du modèle IGa qui maximise la probabilité postérieure sur différentes échelles de temps.....	137
Tableau 3. 8 Estimation des paramètres à partir de LN qui maximise la probabilité postérieure sur différentes échelles de temps.....	138
Tableau 3. 9 Test p-value avec la méthode Kolmogorov-Smirnov (KS). Deux modèles (Modèle 1 & Modèle 2) sont testés par rapport aux données empiriques \mathcal{D}_n . Lorsque p-value $> 0,05$, nous ne rejetons pas l'hypothèse nulle.	142

Tableau 3. 10 Paramètres estimés par maximisation de la vraisemblance des différents modèles sur l'échelle temporelle « minute ». 146

Tableau 3. 11 Comparaison des différents modèles avec la méthode de la somme des erreurs au carré pour l'année 2016 et selon différentes périodes. Le modèle de distribution qui correspond aux données simulées avec le SSE le plus faible est préféré. En 2016, le modèle LN fournit la meilleure estimation des données observées sur les échelles « minute » et « hebdomadaire ». Le modèle IGa fournit meilleure estimation des données observées sur l'échelle « journalière » et le modèle Ga fournit la meilleure estimation des données observées sur l'échelle « hebdomadaire ». 147

Table des matieres

SOMMAIRE.....	5
INTRODUCTION.....	6
CHAPITRE 1 : SYSTÈME COMPLEXE.....	13
1.1. INTRODUCTION	13
1.2. PROPRIÉTÉS ET CARACTÉRISTIQUES D'UN SYSTÈME COMPLEXE.....	16
1.2.1. <i>Processus non-linéaire</i>	16
1.2.2. <i>Processus auto-organisateur et auto-adaptatif</i>	18
1.2.3. <i>Processus ordonné et robuste</i>	21
1.2.4. <i>Phénomène d'Émergence et échelle spatio-temporelle</i>	24
1.3. LE COMPORTEMENT FRACTAL DES SYSTÈMES COMPLEXES	27
1.4. DIFFÉRENCES ENTRE COMPLEXITÉ STATISTIQUE ET COMPLEXITÉ DÉTERMINISTE	31
1.5. CONCLUSION	36
CHAPITRE 2 : MESURE DE LA COMPLEXITÉ : UNE APPROCHE STATISTIQUE. 39	
2.1 INTRODUCTION	39
2.2 LES DIMENSIONS FRACTALES.....	41
2.2.1 <i>Dimension et mesure de Hausdorff</i>	42
2.2.2 <i>Dimension de Capacité (Box Counting)</i>	44
2.2.3 <i>Dimension de Corrélation</i>	45
2.3 DYNAMIQUE SYMBOLIQUE ET THÉORIE DE L'INFORMATION DE SHANNON.....	47
2.3.1 <i>Dimension de Shannon</i> :.....	47
2.3.2 <i>Dynamique Symbolique</i>	49
2.3.3 <i>Information et Entropie de Shannon</i> :.....	51
2.4 DIMENSION ET ENTROPIE GÉNÉRALISÉE.....	55
2.4.1 <i>Dimension de Rényi</i>	55
2.4.2 <i>Spectre de singularité</i>	58
2.4.3 <i>Entropie Généralisée (Rényi & Tsallis)</i> :.....	64
2.5 ANALYSE ET MESURE DE LA COMPLEXITÉ D'UNE SÉRIE ÉCONOMIQUE AU TRAVERS DE LA DYNAMIQUE SYMBOLIQUE ET DE LA THÉORIE DE L'INFORMATION	66
2.5.1 <i>Dynamique Symbolique et Partitions : méthodologie</i>	67
2.5.2 <i>Preuve de présence de mémoire et de prédictibilité dans la série chronologique étudiée.</i> 72	
2.5.3 <i>Entropie et mesure de la complexité</i>	86
2.6 CONCLUSION	93

CHAPITRE 3 : MODÉLISATION DE LA COMPLEXITÉ : APPROCHE SUPER STATISTIQUE.....	96
3.1 INTRODUCTION	96
3.2 SUPER-STATISTIQUES : PRINCIPES FONDATEURS	98
3.2.1 <i>Maxwell-Boltzmann statistique</i>	98
3.2.2 <i>Généralisation de la statistique de Boltzmann</i>	105
3.2.3 <i>Description des familles fondamentales de super-statistiques</i>	108
3.3 SUPERPOSITION DE STATISTIQUES DES SÉRIES FINANCIÈRES : EXPÉRIENCES NUMÉRIQUES	111
3.3.1 <i>Théorie d'Osborne concernant le mouvement Brownien des cours boursiers</i>	112
3.3.2 <i>Expérience numérique</i>	115
3.3.2.1 Système conservatif	115
3.3.2.2 Système dissipatif	119
3.3.3 <i>Volatilité stochastique et super-statistique</i>	125
3.4 ÉTUDE INTRA-JOURNALIÈRE DE LA COMPLEXITÉ DES SÉRIES FINANCIÈRES : COMPARAISON DE SUPER-STATISTIQUES PAR INFÉRENCE BAYÉSIENNE	128
3.4.1 <i>Analyse de données</i>	128
3.4.2 <i>Calibration des modèles : Méthode « Metropolis-Hasting Sampler »</i>	135
3.4.3 <i>Comparaison de modèles via facteurs de Bayes</i>	139
3.4.4 <i>Capacité des modèles IGα et LN à simuler les rendements logarithmiques</i>	141
3.5 PREUVE D'UNE TRANSITION DE STATISTIQUES DE LA VOLATILITE STOCHASTIQUE	144
3.6 CONCLUSION	148
CONCLUSION	151
ANNEXE A	158
ANNEXE B.....	162
ANNEXE C	164
ANNEXE D	167
ANNEXE E.....	168
ANNEXE F.....	169
RÉFÉRENCES	174
TABLE DES FIGURES	183
Liste de tableaux.....	188
TABLE DES MATIERES.....	190