École des Hautes Études en Sciences Sociales

École doctorale de l'EHESS

Centre d'Analyse et de Mathématique sociales

Doctorat

Discipline : Mathématiques et sciences sociales

José MORAN

Statistical physics and anomalous macroeconomic fluctuations

Thèse dirigée par: Jean-Pierre Nadal
Jean-Philippe Bouchaud

Date de soutenance : le 13 Octobre 2020

Rapporteurs
Reimer Kühn, King's College London
Marco Tarzia, Paris-Sorbonne Université

Jury
1 Doyne Farmer, University of Oxford
2 Antoine Mandel, Paris School of Economics
3 Isabelle Méjean, CREST-École polytechnique

STATISTICAL PHYSICS AND ANOMALOUS MACROECONOMIC FLUCTUATIONS

PHYSIQUE STATISTIQUE ET FLUCTUATIONS MACROÉCONOMIQUES ANORMALES

JOSÉ MORAN

REMERCIEMENTS

Je tiens à exprimer mes remerciements à Jean-Philippe Bouchaud, qui pendant ces trois années de thèse m'a toujours proposé des sujets passionnants, en me conférant une grande autonomie sous un encadrement bienveillant. Je remercie également Jean-Pierre Nadal, qui a toujours été disponible pour des conseils sur les projets que je menais.

J'ai également eu l'opportunité de participer aux débuts de la chaire *Econophysix* de l'École polytechnique. Grâce à Michael Benzaquen, j'ai pu participer à des projets et à des missions d'enseignements que j'ai grandement appréciés et je lui en suis très reconnaissant. Merci également aux membres de la chaire pour l'excellente ambiance, notamment à Antoine, Théo, Armine, Pierre-Philippe, Pierre et Mehdi.

Le caractère transdisciplinaire des sujets que j'ai étudiés m'a aussi permis de travailler avec des chercheurs avec qui j'ai pu, en sus de certains projets, avoir des discussions très intéressantes sur l'économie. Il a ainsi été un plaisir de collaborer avec Alan Kirman et Angelo Secchi pendant ma thèse.

Je voudrais également remercier les chercheurs qui ont accepté de faire partie du jury de cette thèse, et en particulier Marco Tarzia et Reimer Kühn pour avoir accepter d'en être les rapporteurs.

La rédaction de ce document a été un travail long mais agréable, les personnes qui ont participé à sa relecture ont été d'une grande aide, merci donc à Joséphine, Théo, Camille, Dhruv et Flora.

Ces trois dernières années se sont passées dans un cadre excellent, et il a toujours été un plaisir de travailler dans les différentes structures qui m'ont accueilli. Je souhaite exprimer ma gratitude à mon laboratoire, le CAMS, ainsi qu'à tous ses membres; j'y ai découvert une ambiance très riche, avec des conférences dans des sujets très intéressants, souvent très différents des miens. Le Laboratoire de Physique de l'École Normale Supérieure a également été un lieu de travail très agréable, qui m'a permis de rester proche de la recherche en physique et de retrouver des personnes très intéressantes que je salue amicalement, je pense notamment à Alexandre, Tristan, Arnaud et les deux Clément. Enfin je voudrais remercier Capital Fund Management, où j'ai passé l'essentiel de mon temps, ainsi que les nombreuses personnes avec qui j'y ai pu interagir, en particulier à Élise, Valentin, Léonard, Amjad, Côme et Philémon. Merci aussi à l'École polytechnique et au Ladhyx, qui m'ont accueilli pendant les derniers mois de ma thèse.

Merci également aux camarades de promotion avec qui je suis resté en contact pendant mes années de doctorat, pour leur bonne humeur et leur soutien. Je pense ainsi aux membres du *Phlysle*, à Geoffrey, Matthieu, Augustin, Camille et en particulier à Dhruv pour les nombreuses discussions que nous avons eues en physique et en économie.

Je tiens aussi à exprimer ma reconnaissance à Daniela et son fauve Nicanor, qui m'ont beaucoup apporté.

Ces dernières années à Paris n'auraient pas été aussi plaisantes sans les nombreuses soirées d'escrime et les compétitions avec l'ECE. Merci à Arthur, Thibaut, Flora, Marion, Adrien, Marguerite, Thaïs, Denis, Yoko, Théo et tous ceux qui m'y ont accompagné, ainsi qu'à Romain pour ses conseils de très grande valeur, à la fois pour l'épée et pour le monde de la recherche.

A lo largo de mis estudios, desde la infancia hasta el doctorado que estoy por culminar, he estado acompañado por el mismo grupo de amigos. Le dirigo, pues, un saludo de agradecimiento a quienes me acompañan de este lado del Atlántico, con quienes he tenido la suerte de convivir estos ultimos años.

Esta página no estaría completa sin un agradecimiento a mis padres, a mi hermana, al resto de mi familia y en particular a mis abuelas, cuyo apoyo, e incluso cuya abnegación, ha sido invaluable.

RÉSUMÉ

Comment des perturbations à l'échelle de l'individu peuvent-elle devenir des grandes fluctuations à l'échelle de l'économie ? Cette question, malgré son étude depuis des décennies, est encore ouverte. Dans cet ouvrage, j'étudie cette énigme à l'aide de méthodes issues de la physique statistique. En partant d'une analyse approfondie des distributions en loi de puissance, je montre qu'une compréhension claire de leur origine et de leurs propriétés permet de mieux appréhender leurs conséquences socioéconomiques. Je propose ensuite un modèle d'économie en réseau, où les entreprises dépendent les unes des autres pour produire, de sorte que la nature même de leurs interactions peut les rendre susceptibles d'amplifier des fluctuations. Ensuite, je soumets au lecteur une étude empirique des propriétés statistiques des taux de croissance et fournis un cadre permettant d'étudier leur dynamique. Dans la partie finale je m'intéresse à des modèles qui exhibent des phénomènes collectifs non-triviaux parce qu'ils considèrent des effets d'imitation ou de mémoire dans les décisions prises par les individus, montrant bien la nécessité de tenir compte de la possible complexité des différentes parties constitutives des modèles économiques.

Mots clef : lois de puissance, économies en réseau, crises, systèmes complexes, phénomènes collectifs, modèles d'agents, taux de croissance d'entreprises.

ABSTRACT

How do microscopic perturbations at the level of an individual grow to become macroscopic fluctuations of the whole economy? Despite decades of effort, this puzzle remains open. In this work, I tackle this problem using methods and techniques from statistical physics. Beginning with a thorough analysis of power law distributions, I argue that understanding their origin and properties helps in elucidating their socio-economic consequences. I then explore a model of an economy where firms interact through a *production network* in a way that causes them to be intrinsically prone to amplify fluctuations. Later on, I conduct an empirical survey of the statistical properties of firm growth rates and provide a framework to study their dynamics. I finally move onto models where non-trivial collective phenomena arise from imitation and memory effects at the level of the individual, highlighting the need of accounting for complexity in economic modelling.

Keywords: power laws, network economies, crises, complex systems, collective phenomena, agent-based models, firm growth rates.

PUBLICATIONS AND PREPRINTS

[1] José Moran, Antoine Fosset, Michael Benzaquen, and Jean-Philippe Bouchaud. "Schrödinger's ants: A continuous description of Kirman's recruitment model". In: *Journal of Physics: Complexity* (June 2020). DOI: 10.1088/2632-072x/aba115. URL: https://doi.org/10.1088/2632-072x/aba115.

[2] José Moran and Jean-Philippe Bouchaud. "Greedy algorithms and Zipf laws". In: *Journal of Statistical Mechanics: Theory and Experiment* 2018.4 (2018). arXiv: 1801.05279, p. 043402. ISSN: 1742-5468. DOI: 10.1088/1742-5468/aab50a.

[3] José Moran and Jean-Philippe Bouchaud. "May's Instability in Large Economies". In: *Physical Review E* 100.3 (2019). arXiv: 1901.09629, p. 032307. ISSN: 2470-0045, 2470-0053. DOI: 10.1103/PhysRevE.100.032307.

[4] José Moran, Antoine Fosset, Davide Luzzati, Jean-Philippe Bouchaud, and Michael Benzaquen. "By force of habit: Self-trapping in a dynamical utility landscape". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30.5 (2020), p. 053123. ISSN: 1054-1500. DOI: 10.1063/5.0009518.

CONTENTS

LIST OF TABLES

INTRODUCTION ET RÉSUMÉ EN FRANÇAIS

a

INTRODUCTION

La *physique statistique* et la *macroéconomie* sont deux disciplines qui n'ont à priori rien à voir l'une avec l'autre. Pour la plupart des personnes que j'ai rencontrées, le mot *physique* évoque plutôt l'étude de corps célestes, de particules subatomiques, de lasers ainsi que d'autres sujets similaires. À première vue, ces notions semblent bien éloignées de l'inflation, des taux de change et de bien d'autres que l'on estime cantonnées au champ d'études de l'économie, et il m'arrive souvent de devoir expliquer comment ces deux disciplines peuvent interagir. Un début de réponse consiste alors à clarifier ce qu'est la physique *statistique*, discipline qui essaie de comprendre le comportement collectif d'une grande quantité d'entités en interaction et qui trouve ses origines dans la théorie cinétique des gaz de Maxwell et Boltzmann. Depuis sa création, ce champ a permis la description précise de phénomènes allant de la transition de phase paramagnétique/ferromagnétique à la dynamique de vol d'un essaim d'étourneaux.

L'origine étymologique du mot *statistique*, qui trahit le fait que cette discipline s'intéressait dans ses débuts à des affaires en rapport avec l'*État*, cache également une curiosité historique. Maxwell a en effet affirmé que sa théorie des gaz devait beaucoup aux travaux d'Adolphe Quetelet [221], un homme fasciné par les régularités des données décrivant la société et également fondateur d'une discipline qu'il a voulu appeler "physique sociale" [54]. Il apparaît alors que c'est la physique statistique qui a d'abord été influencée par la volonté de comprendre des phénomènes collectifs *dans la société*, plutôt que l'inverse. Le nouveau champ des *Systèmes Complexes*, et notamment ce qu'on appelle maintenant l'"*Éconophysique*", a donc une origine toute naturelle quand on s'intéresse à l'histoire de la physique statistique. Il n'est donc pas surprenant que les avancées récentes de la physique statistique puissent être utiles dans l'étude de phénomènes socio-économiques.

C'est dans cette optique, et d'abord en tant que physicien, que j'ai voulu étudier les origines des *fluctuations macroéconomiques anormales*, par opposition aux fluctuations gaussiennes, ou *normales*. La question centrale de mon étude, qui reviendra fréquemment dans cet ouvrage, est de comprendre comment les différentes parties de l'économie, c'est-à-dire les agents individuels mais aussi les entreprises, peuvent s'agréger en un tout dont les différentes fluctuations ne se compensent pas. Cette question, malgré sa concision, représente évidemment un programme de recherche très vaste ; cette thèse vise à étudier quelques-unes des pistes qui pourraient répondre à cet énigme.

Je commence ainsi par l'étude des distributions en loi de puissance. Ces distributions statistiques sont omniprésentes dans les données socio-

économiques, et apparaissent même dans l'étude du langage, notamment lorsque l'on étudie la fréquence d'apparition des mots dans un texte. La première partie de ma thèse est donc consacrée à l'étude de ces distributions, à leurs propriétés et aux types de données dans lesquelles elles apparaissent. En particulier, ces distributions sortent du cadre du théorème central limite habituel, et impliquent donc des phénomènes statistiques *anormaux*. Je présente également une revue rapide de certains des mécanismes qui permettent de comprendre leur origine.

Ces distributions ne permettent malgré tout d'expliquer qu'une partie des fluctuations observées, et il devient nécessaire d'étudier comment les différentes parties de l'économie interagissent à travers les réseaux de production. La deuxième partie de cette thèse étudie ces réseaux, et argumente aussi en faveur de l'idée qu'ils peuvent s'*auto-organiser* dans un état très fragile et susceptible de provoquer de grandes crises à partir de fluctuations microscopiques. Cela introduit l'idée que les fluctuations peuvent être d'origine purement *endogène*, et donc qu'elles surgissent du fait des caractéristiques mêmes de l'économie et des interactions entre ses parties. Cette exploration pourrait conduire à des modèles plus précis, capables non seulement de reproduire certains faits qualitatifs mais aussi d'identifier des mécanismes d'instabilités à partir de données.

Pour ce faire, il est malgré tout nécessaire de comprendre comment analyser des données économiques au niveau de l'entreprise. Dans la troisième partie de cet ouvrage, je soumets au lecteur une étude de données décrivant des entreprises, où je fais une étude de leur croissance. À la lumière du modèle de Gibrat, je propose une étude de la dynamique des taux de croissance de certaines quantités dites fondamentales, comme le chiffre d'affaires d'une entreprise, et montre en quoi elle est différente de la dynamique de variables de marché, comme le prix en bourse d'une action. Cette partie constitue le cœur de mon travail empirique.

Enfin la dernière partie introduit des modèles qui sont des alternatives à l'idée de l'agent rationnel représentatif que l'on trouve fréquemment dans la théorie économique. Ils montrent les effets spectaculaires qui peuvent émerger quand on tient compte des interactions entre agents, ou que l'on va au delà de l'idée que les choix sont faits de manière rationnelle suite à l'optimisation d'une fonction d'utilité statique.

Voulant donner une représentation fidèle de l'approche adoptée pendant mes trois années de thèse, cet ouvrage a été écrit dans la recherche permanente d'un équilibre entre étude de données empiriques, aspects analytiques et simulations numériques. Je pense en effet que les trois sont indispensables pour comprendre des faits empiriques : les modèles devraient avant tout tenter d'expliquer les phénomènes qu'on observe dans les données. Il est clair que les calculs analytiques permettent de comprendre de manière précise et détaillée certains de ces phénomènes, ou de mener des raisonnements logiques qui vont au-delà de ce que nous permet notre langage. Ces raisonnements doivent néanmoins avoir pour point de départ les données elles-mêmes, et il est dangereux d'accorder une valeur de vérité forte à une

théorie mathématique dont les axiomes n'ont pas de fondement sûr. D'un autre côté, les simulations permettent de créer des données *synthétiques* à travers de véritables *expériences numériques* et d'étudier des problèmes qui échappent au traitement analytique à cause de leur complexité. L'étude intelligente de ces simulations peut même justifier une simplification d'un modèle au point de le rendre analytiquement soluble.[1]

C'est avec ces idées en tête que j'ai écrit cette synthèse de mon travail, en essayant d'intégrer les équations au mieux dans le texte et en évitant de leur donner une prédominance face aux données et aux simulations. Je me suis également efforcé de donner des explications intuitives lorsque possible aux phénomènes que j'ai étudié. Mon objectif est avant tout que l'argumentation que je mène soit claire et lisible ; aussi les annexes à la fin de ma thèse fournissent la quasi-totalité du bagage mathématique nécessaire à étayer mes arguments, mais leur lecture n'est pas indispensable.

Les pages suivantes constituent également une présentation résumée de mes travaux qu'un non-spécialiste pourrait lire pour les comprendre sans rentrer dans les détails.

1 Je signale aussi au lecteur que les données (sauf celles protégées par le secret statistique) et les codes utilisés pour générer les figures de cette thèse peuvent lui être fournis sur demande.

b

LOIS DE PUISSANCE

Un fait frappant lorsque l'on examine des données socio-économiques est la présence de ce qu'on appelle les *lois de puissance*. En général, quand on parle d'une loi de puissance reliant deux quantités x et y, on a à l'esprit une relation du type $y \propto x^{\alpha}$ où α est l'exposant de la loi. Cette relation tient sur une plage de valeurs suffisamment large pour x et y. Certaines relations de ce type peuvent sembler assez triviales, comme celle qui relie, par exemple, le côté a d'un cube avec son volume V, c'est-à-dire $V \propto a^{d}$, avec $d = 3$. Il est important de noter que cette relation d'échelle tient toujours si on considère, plutôt qu'un cube, le volume d'une sphère de rayon a ou de tout autre objet tridimensionnel régulier, pour lequel on trouvera toujours la valeur $d = 3$.

Bien que dans le cas cité ci-dessus il existe des raisons physiques évidentes pour cette relation, il est surprenant de constater que des relations similaires existent pour d'autres grandeurs sans qu'on puisse déceler une cause évidente. Un exemple a été trouvé par G.K. Zipf [1] dans ce qui est maintenant connu comme une loi de Zipf. En prenant le roman moderniste de James Joyce *Ulysses*, publié en 1916, il a montré qu'une loi d'échelle émergeait pour la fréquence f_k, ou le nombre d'apparitions, du k-ième mot le plus courant du roman, donné par

$$f_k \propto k^{-\alpha}, \quad \text{with } \alpha \approx 1. \tag{b.1}$$

En outre, il a constaté que cette "loi" s'appliquait à de nombreux textes différents, à peu près indépendamment de la langue dans laquelle ils étaient écrits. En d'autres termes, l'exposant α fluctue d'une langue à l'autre et d'un individu à l'autre, mais la relation donnée en Eq. (b.1) tient toujours. Ce constat, que j'ai voulu montrer dans la Figure b.1 en la calculant pour des livres très différents, a jeté les bases de ce qu'on appelle désormais la linguistique quantitative [73], qui a pour but d'élucider cette régularité et d'autres qui semblent émerger dans le langage naturel.

Une autre "loi", qui a première vue semble sans rapport avec la précédente, a été trouvée en 1886 par Vilfrido Pareto lors de son étude de la répartition des richesses et des revenus en Europe pendant la seconde moitié du 19$^{\text{ème}}$ siècle [208]. En prenant un certain niveau de revenu x, il a compté le nombre de personnes $N_{>}(x)$ déclarant un revenu supérieur à x et a constaté la relation

$$N_{>}(x) \propto x^{-\mu}, \quad \text{avec } \mu \approx 1. \tag{b.2}$$

1 Mais Mandelbrot affirme qu'il a été trouvé plus tôt par Estoup [176, 177].

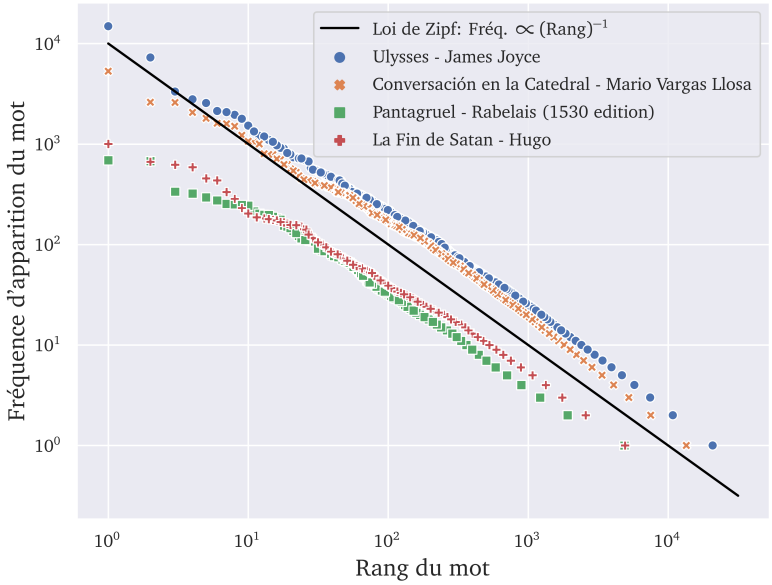


FIGURE b.1 : Relation log-fréquence des mots pour différents textes. D'après cette figure, la fréquence d'apparitions f_k du k-ième mot le plus fréquent dans un texte soit une loi du type $f_k \propto k^{-\alpha}$ avec $\alpha \approx 1$. On peut remarquer les similitudes entre toutes les courbes, bien qu'elles aient été calculées sur des textes aux caractéristiques très différentes, de deux romans quelque peu modernistes (Joyce et Vargas Llosa), à *Gargantua* de Rabelais, écrit en français du seizième siècle, et *La fin de Satan* de Hugo, écrit entièrement en alexandrins.

Comme pour l'observation faite par Zipf pour le langage, cette relation s'étendait sur un large éventail de pays ayant tous des caractéristiques socio-économiques différentes. Bien que l'exposant μ fluctue entre les pays et dans le temps, on constate qu'il est toujours proche de 1. Il convient également de noter que l'exposant décrivant la répartition des revenus est systématiquement plus grand que celui décrivant la richesse [267], ce qui a d'importantes conséquences économiques et sociales.[2] En utilisant les données du livre de Pareto [208], j'ai reproduit ses conclusions dans la Figure b.2 ci-bas pour l'Angleterre et l'Irlande. Comme souligné ci-dessus,

2 En effet, comme le lecteur verra dans le Chapitre 2, cela signifie que la richesse est plus inégalement répartie que les revenus. Les conséquences de cette situation dépassent largement la portée de ma thèse, mais sont bien résumées par ce que Piketty appelle le dilemme de Rastignac. Eugène de Rastignac est un personnage de la *Comédie Humaine* de Balzac, une série de livres qui se déroulent dans la France du 19$^{\text{ème}}$ siècle, qui est convaincu par le machiavélique Vautrin pour arrêter ses études et épouser une riche héritière. L'argument principal de Vautrin, esquissé dans une célèbre tirade, est que l'acquisition d'une fortune permet une existence de rentier, bien meilleure que celle que procure le maigre revenu du travail d'avocat, aussi brillant soit-il. D'un point de vue littéraire, *Le Père Goriot* de Balzac et *L'Argent* de Zola donnent un aperçu intéressant de la tension entre les revenus du travail et ceux du capital. Pour plus d'informations, cette fois-ci d'un point de vue académique, le lecteur intéressé peut consulter les Réfs. [214, 215, 282].

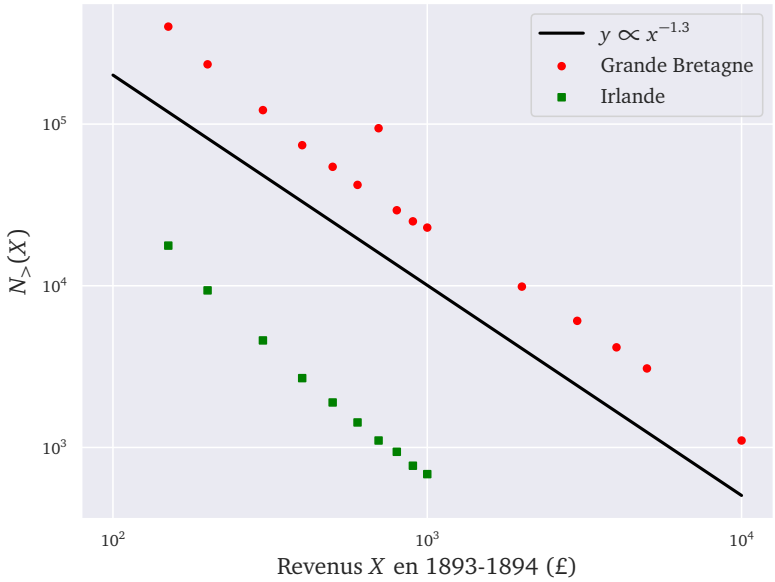


FIGURE b.2 : Nombre d'individus $N_>(X)$ déclarant un revenu supérieur à X à la fin du 19ème siècle en Grande Bretagne (en excluant l'Irlande) et Irlande. Les deux pays satisfont à la relation présentée en Éq. (1.2).

malgré les nombreuses différences entre ces deux pays à l'époque, la relation montrée dans l'Éq. (b.2) tient bon.

On peut naturellement être étonné, puisqu'il n'est pas évident que les deux relations décrites dans les équations (b.2) et (b.1) ont quelque chose à voir l'une avec l'autre, en plus de décrire des quantités très différentes. Pourtant, si l'on prend, pour la clarté, l'exemple du revenu, qu'on classe tous les individus en fonction de leur revenu et qu'on définit $X_{(k)}$ comme le revenu du k-ième individu le plus riche de l'échantillon, on obtient immédiatement que $N_>(X_{(k)}) = k$. A partir de l'équation (1.2) on a alors immédiatement que

$$X_{(k)} \propto k^{-1/\mu}, \tag{b.3}$$

montrant ainsi que les représentations des Figures 1.1 et 1.2, ainsi que les équations (1.1) et (1.2) correspondantes, sont en fait équivalentes en fixant $\alpha = 1/\mu$.

Le nombre de quantités qui suivent des lois de puissance est impressionnant (voir par exemple l'article [88]). C'est le cas de la distribution statistique des degrés [3] dans des réseaux d'interaction de protéines en biologie, de la distribution du degré des serveurs d'internet,[4] du nombre d'appels téléphoniques reçus par les habitants des États-Unis, de l'intensité

3 Ici, le degré indique le nombre d'autres protéines avec lesquelles une protéine donnée interagit.

4 Comme pour les réseaux de protéines, le degré d'un serveur est le nombre d'autres serveurs avec lequel il est connecté.

des guerres définie par le nombre de victimes, du nombre d'espèces par genre biologique chez les mammifères, ou encore de la quantité de citations reçues par les articles scientifiques. De nombreuses analyses existent, par exemple dans [119, 229].

L'omniprésence des lois de puissance dans les phénomènes économiques et sociaux semble donc si forte qu'on ne peut pas s'empêcher de se demander si elles ont vraiment quelque chose de spécial. La première partie de ma thèse présente une vue d'ensemble de ces lois et me permet de motiver la modélisation effectuée dans les parties suivantes. Mon but a été de montrer qu'il y a un intérêt à comprendre leurs conséquences et les façons dont elles émergent.

C'est ainsi que je me suis efforcé de souligner, dans le Chapitre 2 que les quantités réparties selon des lois de puissance conduisent à des phénomènes statistiques contre-intuitifs et que ces phénomènes ont des implications socio-économiques importantes. Enfin dans le Chapitre 3 j'ai voulu expliciter les différentes mécanismes statistiques qui expliquent la présence des lois de puissance, en montrant en particulier que des motifs assez généraux se cachent derrière eux, dans une idée proche de l'*universalité* si chère aux physiciens. Ce dernier point est, à mon avis, un des plus importants dans ce qui a motivé mon travail, car il énonce clairement une des idées clefs de la physique statistique, qui consiste en la recherche de "lois" statistiques qui décrivent le comportement collectif de systèmes qui, à première vue, n'ont rien à voir entre eux. Dans la suite, je présenterai un résumé du Chapitre 2 portant sur les propriétés statistiques des lois de puissance et du Chapitre 3 qui décrit des mécanismes qui permettent de les générer.

b.1 PROPRIÉTÉS STATISTIQUES DES LOIS DE PUISSANCE

b.1.1 *Quelques définitions*

Avant de s'attaquer à la suite, il convient d'abord de définir ce que j'entends par une distribution statistique en loi de puissance. Comme je suis avant tout motivé par la description de données empiriques, et pas par la définition d'une distribution de probabilités en tant qu'objet mathématique abstrait, je propose la définition suivante.

Face à un ensemble d'observations $(X_1, \dots, X_N)$ de, par exemple, la taille de villes ou les revenus d'individus, on dit que la quantité X est distribuée selon une loi de puissance si la relation (b.2) tient pour $X > X^+$ avec un seuil $X^+ \gg 1$. Lorsque le nombre d'observations N est important, il est plus facile de prendre une limite continue et de considérer la quantité $P(x)$, définie par

$$P(x) = \lim_{N\to\infty} \lim_{\Delta x\to 0} \frac{N_>(x+\Delta x) - N_>(x)}{N}, \tag{b.4}$$

où N est le nombre total d'individus dans l'échantillon. Cette quantité est assimilée à une densité de probabilité.

On peut alors aussi définir la fonction de répartition complémentaire (cdf complémentaire, de *complementary distribution function* en anglais), ou fonction de survie, comme :

$$P_>(x) := \int_x^\infty \mathrm{d}s\ P(s), \tag{b.5}$$

qui satisfait par ailleurs $P_>(x) \approx N_>(x)/N$. Cette fonction peut être définie avec la fonction de répartition $P_<(x)$ à travers la relation $P_>(x) = 1 - P_<(x)$.

$P(x)$ est donc la densité de probabilité associée à X, qui est donc modélisée comme une variable aléatoire. En conséquence de l'équation (b.2), cette densité satisfait à $P(x) \propto x^{-1-\mu}$, ou de manière équivalente à $P_>(x) \propto x^{-\mu}$ sur une plage de valeurs $x > X^+$. Si cette relation tient sur l'ensemble des valeurs observées $[x_{\min}; +\infty[$ alors on a affaire à une loi de puissance pure, dite aussi distribution de Pareto, soit

$$P(x) = \frac{\mu x_{\min}^{\mu}}{x^{1+\mu}}. \tag{b.6}$$

Le cadre des lois de puissance pure est néanmoins trop contraignant pour l'étude de toutes les distributions que l'on dit être des lois de puissance. En particulier, une loi de puissance pure impliquerait une ligne parfaitement droite dans les Figures b.1 et b.2. En conséquence, je définis une loi de puissance comme distribution telle que $P(x|x > X^+)$, c'est-à-dire la densité de probabilité de S conditionnelle à être supérieure à X^+, est raisonnablement bien approchée par une densité de Pareto comme dans Eq. (2.3) pour un certaine valeur de l'exposant μ. Cela revient à prendre $x_{\min} = X^+$ dans l'équation (b.6). Cela équivaut à dire que le comportement en loi de puissance est défini dans les queues de la distribution seulement, et donc que X est distribué selon une loi de puissance si sa densité vérifie

$$P(x) \underset{x \gg 1}{\sim} x^{-1-\mu}. \tag{b.7}$$

Naturellement, bien que j'aie introduit des définitions pour une variable aléatoire non négative $X > 0$, il est possible de les étendre à toute variable aléatoire à valeurs réelles X, en se laissant la possibilité d'avoir deux exposants μ_+ et μ_- décrivant les queues droite et gauche. Ainsi la densité doit vérifier asymptotiquement

$$P(x) \approx \begin{cases} A_+ x^{-1-\mu_+} & \text{for} \quad x \gg 1 \\ \\ A\_x^{-1-\mu_-} & \text{for} \quad x \ll -1 \end{cases} \tag{b.8}$$

avec $A_\pm$ des constantes positives fixant le poids relatif de chaque queue.

Ces définitions étant posées je peux maintenant m'attaquer aux propriétés statistiques des lois de puissance.

b.1.2 *Moments et valeurs extrêmes*

Suivant une tradition très présente dans les références [65] et [66] et en suivant de façon assez fidèle ce qui je présente en section 2.2, j'introduis ici les propriétés statistiques des lois de puissance *via* la théorie des valeurs extrêmes, ce qui a l'avantage de donner de l'intuition aux phénomènes qu'elles décrivent.

En effet, en regardant l'équation (2.4) on constante que les moments de la distribution de X ne sont pas toujours définis. En prenant la quantité $\mathbb{E}[X^m] = \int \mathrm{d}x\ P(x)x^m$, il est immédiat de constater que l'intégrale se comporte comme $x^{m-1-\mu}$ pour x grand, et donc que l'intégrale converge seulement pour $m > \mu$. [5] En conséquence, l'écart type ou la variance ne sont pas définis pour une loi de puissance avec $\mu < 2$, et la moyenne n'est pas définie non plus lorsque $\mu < 1$.

Pour un jeu de données réel avec un nombre fini d'observations N, les moments empiriques sont toujours finis, et leur divergence doit donc être interprétée dans ce contexte. Ayant un ensemble d'observations N $(X_1, \dots X_N)$, on s'attendrait, en raison de la loi des grands nombres, à la convergence :

$$\frac{1}{N}\sum_{i=1}^{N} X_i^m \underset{N\to\infty}{\longrightarrow} \mathbb{E}[X^m]. \tag{b.9}$$

Or la discussion précédente implique que l'expression de gauche diverge lorsque $N \to \infty$ si $m \leq \mu$. En prenant par exemple $\mu < 1$ et en examinant le comportement de la moyenne empirique, on peut en fait gagner en intuition en remarquant que la valeur maximale de l'échantillon, dénommée M_N, devrait satisfaire $P_>(M_N) \approx 1/N$. En utilisant Eq. (b.5) on a alors immédiatement que

$$M_N \sim N^{\frac{1}{\mu}}, \tag{b.10}$$

et donc que le rapport M_N/N se comporte comme $N^{\frac{1-\mu}{\mu}}$ et diverge donc quand $N \to \infty$. L'expression à gauche dans l'équation (b.9) est donc nécessairement divergente, mais on est maintenant capables d'énoncer un résultat plus fort : la somme des variables X_i est faite de N termes, dont le plus grand a un comportement en $N^{\frac{1}{\mu}}$, ce qui implique qu'il domine tous les autres. J'y reviendrai lorsque j'examinerai les répercussions des lois du pouvoir sur l'inégalité. Le même raisonnement s'applique à des moments comme X^m en notant que la variable $Y = X^m$ suit une loi de puissance de densité $P(y) \propto y^{-1-\frac{\mu}{m}}$.

5 En fait, cela permet aussi de donner une définition alternative pour une loi de puissance : on définit alors μ comme l'infimum de $\{m|\mathbb{E}[S^m] < \infty\}$, la distribution est une loi de puissance si $\mu < \infty$ et on a directement une définition de μ. Cette définition indique également qu'une distribution avec une densité qui se comporte asymptotiquement comme $P(x) \approx \ln(x)^\alpha x^{-1-\mu}$ est assimilée à une loi de puissance. En pratique, il est très difficile de distinguer ce comportement logarithmique lors d'une analyse empirique.

Ce résultat est à comparer avec ce qu'on obtiendrait avec d'autres distributions, comme une loi exponentielle. Prenons alors une variable aléatoire X de densité $P(x) = e^{-x}$, et donc avec $P_>(x) = e^{-x}$. Avec le même raisonnement que celui donné plus haut, on constate que la plus grande valeur parmi N observations doit se comporter comme $M_N \sim \ln(N)$. Dans ce cas, $M_N/N \to 0$ dans la limite de N grand, et donc cette valeur maximale ne représente qu'une fraction évanescente de la somme de tout l'échantillon.

Il est en fait possible d'aller plus loin et de définir des "classes d'universalité" pour les maxima de variables aléatoires. Si on considère, par exemple, N observations d'une variable dont la queue se comporte en

$$P_>(x) \approx \frac{A^\mu}{x^\mu}, \tag{b.11}$$

alors il est possible de définir une variable $u = \frac{M_N}{AN^{1/\mu}}$, qui est donc d'ordre 1 avec une probabilité élevée, et de montrer qu'elle converge en loi vers une loi de Fréchet de densité

$$P(u) = \frac{\mu}{u^{1+\mu}} e^{-u^{-\mu}}. \tag{b.12}$$

Il convient alors de remarquer que la loi de Fréchet est elle-même une loi de puissance, de même exposant que celle de X. Le passage au maximum préserve alors la structure en loi de puissance.

Cette universalité du maximum existe aussi pour des lois dont les queues décroissent plus vite. En prenant un exemple donné dans [65] on peut prendre une variable aléatoire de densité

$$P(x) \approx \frac{A}{x^\alpha} \exp\left(-Bx^\delta\right), \quad B, \delta > 0, \tag{b.13}$$

et donc de fonction de survie donnée par

$$P_>(x) \approx \frac{\delta A}{B} x^{1-\delta} \exp\left(-Bx^\delta\right). \tag{b.14}$$

Le maximum est alors donné par $M_N \sim \left(\frac{\ln(N)}{B}\right)^{1/\delta}$, et la variable u définie par l'équation

$$M_N = \left(\frac{\ln(N)}{B}\right)^{1/\delta} + \left(\frac{\ln(N)}{B}\right)^{-1/\delta} \frac{u}{B\delta}, \tag{b.15}$$

est distribuée selon la loi dite de Gumbel,

$$P(u) = \exp(-u - \exp(-u)), \tag{b.16}$$

et a donc des queues exponentielles. Ce point est clef pour comprendre l'intuition derrière la condensation du *Random Energy Model* (REM) donnée en section 3.3.3.1 et par la suite la condensation dans les modèles multiplicatifs, ce qui a des conséquences importantes dans les phénomènes d'inégalités de richesse, pour ne citer qu'un exemple.

b.1.3 *Sommes de variables aléatoires*

Considérons des sommes de la forme

$$S_N = \sum_{i=1}^{N} X_i, \tag{b.17}$$

de telle sorte qu'on est tenté d'invoquer le théorème central limite et d'écrire que

$$S_N = N\mathbb{E}[X] + \sqrt{N\mathbb{V}[X]}\xi, \tag{b.18}$$

avec ξ une variable gaussienne centrée réduite. Cette écriture présuppose l'existence des moments $\mathbb{E}[X]$ et $\mathbb{E}\left[X^2\right]$, ce qui veut dire que le théorème central limite "standard" ne fonctionne plus pour des lois de puissance avec $\mu \leq 2$.

Il est néanmoins possible d'étendre le théorème central limite à ces variables, en écrivant le théorème central limite généralisé dont la preuve est esquissée dans l'Appendice A.

Prenant alors une variable X de densité $P(x) \approx Ax^{-1-\mu}$ il vient que

$$S_N = \begin{cases} N\mathbb{E}[X] + (AN)^{1/\mu}\,\xi_G & \text{for} \quad \mu > 2 \\ N\mathbb{E}[X] + (AN)^{1/\mu}\,\xi_L & \text{for} \quad 1 < \mu \leq 2\,, \\ (AN)^{1/\mu}\,\xi_L & \text{for} \quad \mu < 1 \end{cases} \tag{b.19}$$

où ξ_G est une variable aléatoire gaussienne et ξ_L une variable de Lévy de paramètre μ. On peut comprendre les différents comportements, en $\sqrt{N}$ dans l'équation (b.18) et en $N^{1/\mu}$ dans l'autre cas, en raisonnant sur les fluctuations typiques de la moyenne de S_N : si la variance est définie, les fluctuations se comportent comme dans le cas gaussien et sont en $\sqrt{N}$, alors que dans le cas en loi de puissance ces fluctuations doivent être imputées au maximum des variables X_i, qui se comporte comme $N^{1/\mu}$.

Ce raisonnement nous permet aussi de comprendre le comportement des distributions de Lévy, qui sont aussi des lois de puissance, puisque leur densité vérifie asymptotiquement

$$L(\xi) \approx \frac{\mu A}{|\xi|^{1+\mu}}. \tag{b.20}$$

b.1.4 *Phénomènes de condensation*

Il est parfois intéressant aussi de calculer des sommes de moments d'une variable X, ou, de façon équivalente, de calculer des normes q d'un échantillon de N observations. On étudie alors des objets tels que

$$S_{N,q} = \left(\sum_{i=1}^{N} X_i^q\right)^{1/q}. \tag{b.21}$$

Pour une variable en loi de puissance, il est immédiat que

$$S_{N,q} \approx X_{\max}, \quad \text{for } q > \mu, \tag{b.22}$$

dans une situation qui rappelle les transitions de phase en physique statistique, où q joue ici le rôle d'une température inverse et $S_{N,q}$ celui d'une fonction de partition.

Ceci est à mettre en lien avec l'indice dit de herfindahl utilisé en économie [137], qu'on peut également identifier à l'*Inverse Participation Ratio* utilisé en physique théorique de la matière condensée [270, 41]. Ces quantités reviendront dans les Chapitres 5, 6 et 8

L'indice de herfindahl associé aux variables $X_1, \dots, X_N$ est défini par

$$H = \frac{\sum_{i=1}^N X_i^2}{\left(\sum_{i=1}^N X_i\right)^2} = \sum_{i=1}^N \left(\frac{X_i}{\sum_{j=1}^N X_j} \right)^2, \tag{b.23}$$

et il est d'ordre 1 for $\mu \lessapprox 2$. En revanche, pour des variables X_i qui sont toutes de taille équivalente, on doit trouver $H \approx 1/N$.

b.1.5 *Impact dans les inégalités de richesse*

Considérons le cas où les variables X décrivent la richesse d'individus, avec X_i la richesse de l'individu i, et prenons les comme étant distribuées selon une loi de puissance pure.

Dans ce cadre, la richesse totale de la société $W_N = \sum_{i=1}^N X_i$. On peut alors se demander comment la richesse totale est répartie entre les N individus. Comme indiqué plus haut, la question est triviale lorsque $\mu < 1$ car la richesse de l'individu le plus riche est de l'ordre de $N^{1/\mu} \gg N$, ce qui implique que pratiquement toute la richesse est détenue par une poignée d'individus.

Quand $\mu > 1$, on peut alors calculer explicitement la part de la richesse $w(q)$ détenue par la fraction $q \in [0;1]$ des individus les plus pauvres. Elle est alors donnée par

$$w(q) := \frac{N \int_1^{x_q} \mathrm{d}x \; p(x) x}{W_N} = 1 - (1-q)^{\frac{\mu-1}{\mu}}. \tag{b.24}$$

On peut alors définir l'indice de Gini, utilisé dans la littérature économique pour quantifier le niveau d'inégalité au sein d'un pays, comme l'aire entre la courbe $(q, w(q))$, également connue sous le nom de courbe de Lorenz, et la courbe (q,q). Un calcul direct donne l'indice de Gini en fonction de l'exposant de queue μ,

$$G(\mu) = \frac{2\mu}{2\mu - 1} - 1, \tag{b.25}$$

et on remarque alors qu'il tend vers 0 lorsque $\mu \to \infty$, comme on s'y attend. Une illustration de ce calcul est fournie en Figure b.3.

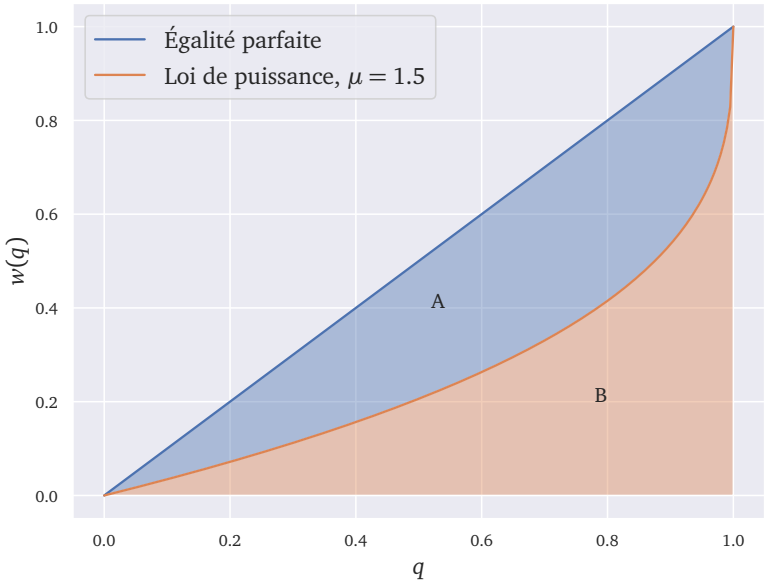


FIGURE b.3 : Une illustration de l'indice de Gini. L'indice de Gini est défini comme le rapport d'aires $G = A/(A+B)$. L'axe des abscisses décrit le quantile q tandis que l'axe des ordonnées donne la fraction de la richesse totale $w(q)$ détenue par ce quantile. La courbe supérieure correspond à $w(q) = q$, une situation d'égalité parfaite, tandis que la courbe inférieure correspond à une situation où la richesse est répartie selon une loi de puissance avec $\mu = 1.5$.

Une autre situation intéressante dans ce contexte est connue sous le nom de "règle des 80/20", qui stipule, par exemple, que les 20% les plus riches possèdent 80% de la richesse, ou que les 20% mots les plus couramment utilisés dans un texte représentent 80% de son contenu. Cela équivaut à dire que $w(0.8) = 0.2$, et on peut facilement trouver l'exposant de queue pertinent $\mu = \frac{\ln(5)}{\ln(4)} \approx .1.16$, très proche des exposants trouvés dans les données empiriques.

Tout cela montre la pertinence des lois de puissance dans les analyses statistiques de données en sciences économiques et sociales. Il faut néanmoins tenter d'apporter des explications à leur émergence.

b.2 MODÈLES POUR DES LOIS DE PUISSANCE

La plupart des modèles présentés dans le Chapitre 3 partent de modèles de croissance multiplicative stochastique. Ces derniers, munis d'hypothèses supplémentaires assez faibles, conduisent à des lois de puissance de manière très naturelle. Pour ces modèles, l'exposant μ dépend de différents paramètres décrivant le processus. Il en va de même pour le temps caractéristique de convergence vers la distribution stationnaire. Le cas Zipf $\mu \approx 1$

apparaît souvent comme un cas limite, pour lequel ce temps est divergent. Les modèles multiplicatifs sont présentés en Section 3.3.

Pour une autre classe de modèles, dits *critiques*, l'exposant ne dépend pas des détails microscopiques du processus génératifs. Néanmoins le comportement en loi de puissance pure n'est retrouvé que sur un *point critique*, analogue des points de transition de phase en physique statistique, comme la transition entre eau solide et liquide. Il est toutefois possible de formuler des arguments pour dire que la dynamique du système le pousse naturellement vers ce point critique, à l'image d'un tas de sable sur lequel on ajoute des grains et qui croît vers une situation instable où il s'effondre avec des avalanches, avant de recommencer à grandir. Ces modèles sont abordés en détail dans la Section 3.4.

Je décris ci-dessous de façon résumée les résultats présentés dans le Chapitre 3, mais les preuves et les descriptions précises des modèles sont faites dans le corpus du texte en anglais. Je note également que, dans le but d'insister sur le caractère "universel" de ces mécanismes, j'ai choisi de commencer par des modèles qui décrivent d'abord la présence de lois de puissance dans le langage naturel, avant de montrer que ces mêmes mécanismes sont à l'œuvre dans des modèles qui décrivent des quantités "économiques" comme la richesse ou la population d'une ville.

LE MODÈLE DES SINGES DACTYLOGRAPHES Dans ce modèle, on imagine des singes qui écrivent sur un clavier en choisissant des touches de manière aléatoire. Une lettre quelconque est choisie avec probabilité $1-p$ et la touche espace est choisie avec probabilité p. On trouve alors que la fréquence d'apparition des mots est déterminée par une loi de puissance d'exposant

$$\mu = \left(1 + \frac{\ln(1-p)}{\ln(n)}\right)^{-1}, \tag{b.26}$$

avec n la taille de l'alphabet. On retrouve alors une loi de Zipf $\mu \approx 1$ dans la limite $p \to 0$.

LE MODÈLE DE YULE/SIMON Il s'agit ici d'un modèle qui a une double interprétation. On peut prendre l'interprétation de Simon, dans laquelle on s'imagine en train d'écrire un texte : à chaque pas de temps on choisit un mot existant avec probabilité $1-\alpha$ et on le reprend comme mot suivant dans le texte, ou bien avec la probabilité complémentaire on ajoute un mot qui n'est pas encore présent. L'autre interprétation possible, et qui semble pourtant bien éloignée, est celle de Yule : on imagine des espèces classifiées dans des genres biologiques, et on ajoute des espèces à chaque genre avec un taux proportionnel à s, ou bien on ajoute des genres par des mutations avec un taux g.

On peut alors s'intéresser à la distribution statistique de la fréquence des mots ou des tailles (en termes de nombre d'espèces) des genres biologiques.

Il est alors possible de retrouver des correspondances entre les deux modèles, aboutissant à la même résolution mathématique, et on trouve alors des lois de puissance d'exposant

$$\mu = 1 + \frac{g}{s} = \frac{1}{1-\alpha}. \tag{b.27}$$

La loi de Zipf $\mu = 1$ s'obtient dans la limite $\alpha, g \to 0$, où la probabilité d'ajout de nouvelles entités, qu'il s'agisse de nouveaux mots ou de nouveaux genres, tend vers 0. Il est également possible de formuler le modèle d'attachement préférentiel, dit de Barabási-Albert et modélisant la croissance des réseaux, dans le langage de Yule/Simon dans le cas $g = s$.

LES MODÈLES DE RÉDUCTION DE L'ESPACE DES PHASES il s'agit d'un modèle initialement formulé par Corominas-Murtra, Hanel et Thurner [94] et que j'ai repris pour une résolution dans la limite continue avec Jean-Philippe Bouchaud dans [190]. Dans ce modèle, on imagine une dynamique où l'espace des possibilités se rétrécit au fur et à mesure qu'on avance. On imagine par exemple une dynamique où on a une position à l'instant t donnée par $x(t) \in [0;1]$. La dynamique veut alors qu'on choisisse une nouvelle position $x(t+1) \in [x(t);1]$ uniformément, ou qu'on en choisisse une autre dans $[0;1]$ avec probabilité $\propto \phi$.

Le modèle peut aussi être étendu vers un cas où $x(t)$ représente un "niveau d'énergie" E de distribution, ceux-ci étant distribués selon une densité $\mathscr{P}(E)$. La position dans $[0;1]$ équivaut alors à $\int_E^\infty \mathrm{d}E'\ \mathscr{P}(E')$, et la dynamique veut qu'on choisisse un niveau $E' < E$ avec probabilité $\mathbf{1}(E' < E) / \int_{-\infty}^{E} \mathrm{d}E'\ \mathscr{P}(E')$.

Il est possible de résoudre le modèle exactement, et de montrer que la distribution des populations des niveaux d'énergie à l'équilibre ne dépend pas du choix de $\mathscr{P}$, et qu'elle est donnée par une loi de puissance d'exposant

$$\mu = 1 + \phi, \tag{b.28}$$

et que le temps de convergence vers cet équilibre est $\propto 1/\phi$. Une extension du modèle, où on imagine qu'on choisit un niveau d'énergie ou une position dans $[0;1]$ uniformément parmi *toutes* les possibilités, et qu'on accepte ce choix seulement s'il correspond à un état d'énergie plus basse ou un état tel que $x(t+1) > x(t)$, mène cette fois-ci à une loi de puissance dont l'exposant vaut toujours

$$\mu = 3/2. \tag{b.29}$$

Comme pour le modèle original, j'ai contribué à la solution exacte de cette extension dans [190] en utilisant la méthode des caractéristiques.

LE MODÈLE D'ÉNERGIE ALÉATOIRE DE DERRIDA Dans ce modèle, on considère N variables aléatoires log-normales, que l'on peut écrire comme

$x_i \propto e^{\sigma\sqrt{t}\xi_i}$ avec ξ_i des variables gaussiennes. Ce modèle est utilisé pour décrire de manière effective la physique de certains matériaux désordonnés, comme la limite $p \to \infty$ du *p-spin model*, mais il peut aussi décrire la façon dont se répartit la richesse de N individus avec chacun un capital qui évolue de manière multiplicative.

Il est alors possible, en utilisant la théorie des valeurs extrêmes, de montrer que ce modèle est équivalent à une distribution en loi de puissance d'exposant

$$\mu = \sqrt{\frac{2\ln(N)}{\sigma^2 t}}, \tag{b.30}$$

dans la limite de T et N grands avec μ fixé.

CROISSANCE MULTIPLICATIVE AVEC MORT ET/OU BARRIÈRE Dans les deux cas, il s'agit de l'étude d'un mouvement brownien géométrique avec un *drift* logarithmique m et une log-volatilité σ^2, auquel on ajoute un taux de mortalité $\lambda \geq 0$ qui équivaut à un redémarrage de la position de la particule sur une certaine position. On retrouve alors une loi de puissance exacte dans l'état stationnaire lorsque $\lambda > 0$, avec un exposant

$$\mu = -\frac{m}{\sigma^2} + \frac{\sqrt{m + 2\lambda\sigma^2}}{\sigma^2}. \tag{b.31}$$

Si l'on considère plutôt un modèle sans mort, c'est-à-dire avec $\lambda = 0$, mais avec un seuil inférieur limitant la position $x(t)$, on obtient également une loi puissance exacte lorsque $m < 0$, d'exposant

$$\mu = \frac{|m|}{\sigma^2}. \tag{b.32}$$

LE MODÈLE DE BOUCHAUD-MÉZARD Il s'agit d'un cadre très similaire à celui de croissance multiplicative avec barrière réfléchissante, présenté dans le dernier paragraphe. On modélise des agents dont la richesse croît multiplicativement, mais avec un terme de "redistribution" J, qui modélise des impôts ou des échanges commerciaux et qui a un rôle similaire à la barrière réfléchissante définie ci-dessus. On trouve alors que la richesse des agents, mesurée en unités de la richesse moyenne dans l'économie, est asymptotiquement distribuée comme une loi de puissance d'exposant

$$\mu = 1 + \frac{2J}{\sigma^2}, \tag{b.33}$$

et on retrouve encore une loi de Zipf dans la limite $J \to 0$. Je tiens aussi à remarquer que le modèle ne converge pas lorsque $J = 0$.

PROCESSUS DE BRANCHEMENT DE POISSON ET AVALANCHES Cette famille de modèles donne toujours, en son *point critique*, des avalanches dont la taille est distribuée selon une loi de puissance d'exposant

$$\mu = 3/2. \tag{b.34}$$

Cette valeur est indépendante des détails sous-jacents à chaque modèle, qu'il s'agisse des interprétations de Galton-Watson, Cournot/Bienaymé, Erdös-Renyi ou du modèle d'Ising en champ aléatoire.

C'est très semblable à ce qui se passe dans les modèles présentant une transition de phase du deuxième ordre en physique statistique, comme par exemple le modèle d'Ising en dimension $d > 1$.

On peut néanmoins superposer plusieurs instances ces processus avec des valeurs du paramètre λ proches de sa valeur critique $\lambda_c = 1$; ceci conduit à une distribution d'avalanches qui est donnée par la superposition d'une loi de Zipf avec des avalanches ou des clusters dont la taille est comparable à celle de tout le système.

On constate alors qu'on trouve des lois de puissance dans de nombreux systèmes complexes, et que les mécanismes qui les expliquent ont des propriétés communes. La présentation qui est faite dans cet ouvrage est naturellement incomplète, mais permet, je pense, d'acquérir une certaine intuition à ce sujet. En ce qui concerne les conséquences de l'existence de ces lois statistiques, outre celles qui entourent la question sociétale de l'inégalité des richesses et des revenus qui sont bien expliquées dans le Chapitre 2, je montrerai également qu'elles jouent un rôle dans l'origine des grandes fluctuations en économie.

C

FLUCTUATIONS ÉCONOMIQUES ANORMALES

La persistance de grands écarts par rapport à la tendance dans les séries de données économiques, comme celles qui décrivent le produit intérieur brut (PIB) des pays, est un casse-tête économique vieux de plusieurs décennies. Un exemple très clair de ce phénomène est donné dans la figure c.1, où je montre le taux de croissance annuel glissant du PIB réel des États-Unis depuis 1948. Le graphique montre des fluctuations très importantes. On peut attribuer certaines à de grandes crises, comme le choc pétrolier de 1973, la crise financière mondiale de 2007—2008 et, plus récemment, la crise du Covid-19 de 2020 qui se déroule au moment même ou j'écris ce manuscrit. Au total, le taux de croissance pour cette période est de 3.18%, mais son écart-type est comparable et s'élève à 2.57%.

Une explication naturelle à ce phénomène, qui nous apparaît aujourd'hui comme très naïve, a été proposée au début du 19$^{\text{ème}}$ siècle par Herschel et Carrington, mais aussi, et de façon célèbre, par Jevons en 1875 (voir la Réf. [124] et les références qui y figurent) dans ce qui est maintenant connu sous le nom de "théorie des taches solaires", ou "*sunspot theory*" en anglais.

En bref, ils ainsi suggéré que des fluctuations exogènes, comme celles de l'activité solaire dans le cas de Jevons, pouvaient être à l'origine des variations dans le succès des récoltes et des prix de certains biens, ce qui aurait in fine un impact sur l'ensemble de l'économie. Bien qu'ils aient rencontré beaucoup de scepticisme de leur vivant, leur explication reste présente aujourd'hui dans l'idée de fluctuations exogènes.

Dans le paradigme des fluctuations exogènes, on imagine que l'économie doit être intrinsèquement en équilibre, et que tous les flux matériels et monétaires devraient normalement s'équilibrer ; la seule origine possible des fluctuations est alors la "main de Dieu", c'est-à-dire des chocs totalement exogènes comme les crises politiques ou les catastrophes naturelles qui éloignent temporairement l'économie de son équilibre, avant qu'elle n'y revienne naturellement par des mécanismes d'autorégulation. Ceci constitue le cœur des outils utilisés par les banques centrales pour définir leurs politiques.

En effet, les modèles d'"équilibre général dynamique et stochastique" (dont l'acronyme en anglais est DSGE) considèrent des économies constituées chacune d'une "entreprise représentative" qui englobe toute l'industrie et d'un "ménage représentatif" qui capte le comportement du travail et des consommateurs. Ces deux entités sont alors influencées par des chocs aléatoires [123]. Ces modèles ont été vivement critiqués dans le sillage de la crise financière de 2007—2008, notamment pour leur incapacité à reproduire la grande volatilité observée dans les données réelles et pour la

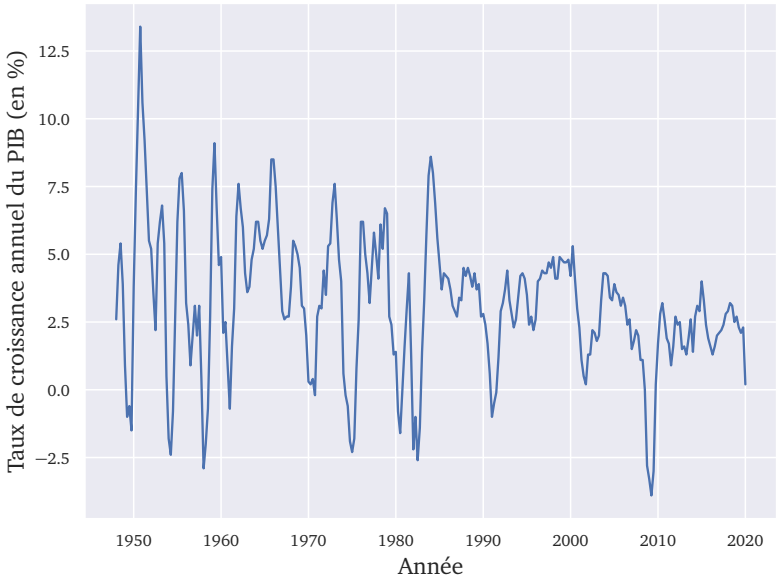


FIGURE c.1 : Taux de croissance annuel du PIB réel américain, corrigé des variations saisonnières et après soustraction de la tendance. Le pourcentage est calculé en fonction de la variation par rapport au même trimestre de l'année précédente. La moyenne depuis 1948 est de 3,18%, mais avec un écart-type de 2,57%.

façon dont ils se sont effondrés lorsqu'ils ont été confrontés à des situations de crise [49, 252].

Comprendre les origines des grandes fluctuations économiques est donc un enjeu crucial, et cela aurait des répercussions directes sur les politiques publiques. Un premier pas pour les comprendre consiste en étudier comment les chocs exogènes sur des agents représentatifs se comportent au niveau agrégé, dans le but ensuite de savoir comment ils peuvent être amplifiés par des fluctuations qui seraient alors d'origine *endogène*.

Prenons un exemple simple pour comprendre la difficulté de ce projet. Si on imagine l'économie étasunienne contemporaine comme étant constituée de N entreprises équivalentes, toutes des copies d'une entreprise "représentative" imaginaire, on peut imaginer que chacune a une production mesurée par une quantité $s_i(t) = s(t)$ à l'instant t, que l'on peut se représenter comme son chiffre d'affaires pour avoir les idées claires. La production totale de l'économie serait alors

$$S(t) = Ns(t), \tag{c.1}$$

et on peut ensuite imaginer que chaque entreprise voit sa production évoluer de manière stochastique, selon une évolution du type

$$s_i(t+1) = (1 + m + \sigma\eta_i(t))\, s_i(t), \tag{c.2}$$

avec les η_i des variables gaussiennes centrées, réduites et indépendantes. On se représente alors chaque entreprise comme ayant un taux de croissance

moyen de m pour-cent par période de temps, avec un écart type donné par σ.

Si nous considérons ensuite le taux de croissance de toute l'économie, celui-ci est donné par

$$\frac{S(t+1)}{S(t)} = \frac{\sum_{i=1}^{N}(1+m+\sigma\eta_i(t))}{N} \approx 1+m+\frac{\sigma}{\sqrt{N}}\eta, \tag{c.3}$$

avec η encore une gaussienne centrée et réduite.

Ce traitement revient à une application naïve du théorème central limite, selon lequel les fluctuations autour de la tendance $1+m$ se comporteraient comme $1/\sqrt{N}$. D'après cet argument on s'attendrait à ce qu'une économie aussi grande que celle des États-Unis, qui compte des millions d'entreprises, ne fluctue pas du tout !

Une réponse possible à cette question a été proposée par Gabaix, à travers ce qu'il appelle la "granularité".

C.1 LA GRANULARITÉ DE GABAIX

L'argument de Gabaix peut être rapidement résumé comme une façon de "briser" le théorème central limite. Il part en effet du constat que la "taille" des entreprises, définie par exemple par leur chiffre d'affaires, est distribuée selon une loi de puissance avec un exposant proche de 2. Ceci est empiriquement vrai pour beaucoup de pays, et ce indépendamment de la période considérée.

Ainsi il n'y a plus d'"entreprise représentative", et on peut dire alors que certaines fluctuations sont dues à des entreprises qui représentent une part non-négligeable de l'économie. À titre d'exemple, on peut reprendre l'exemple de Nokia qui a contribué en 2004 de 1.6 points de pourcentage au PIB Finlandais [203].

En reprenant une croissance comme celle donnée en équation (c.2), on peut dire que la somme des ventes de N entreprises de tailles s_i hétérogènes a un taux de croissance

$$\frac{S(t+1)}{S(t)} = \sum_{i=1}^{N}(1+m+\sigma\eta_i)\frac{s_i(t)}{S(t)}, \tag{c.4}$$

une hypothèse qui peut sembler irréaliste à cause de la possible hétérogénéité des entreprises, notamment au vu des résultats présentés plus loin dans le Chapitre 6. Néanmoins, cette idée tient dans la limite $N \gg 1$ pour des valeurs de m et σ hétérogènes tant que leurs distributions ont une variance finie, ce qui est le cas empiriquement.

En prenant la variance de l'équation précédente, on peut alors dire que

$$\sigma_S^2 = \sigma^2 H, \quad H = \sum_i \left(\frac{s_i(t)}{S(t)}\right)^2, \tag{c.5}$$

où H est le herfindahl, défini en détail dans la Sec. 2.3.1. Lorsque la distribution des variables s_i a une queue lourde, telle que $\mathbb{V}[s_i] = \infty$ ou, de façon équivalente, lorsqu'elle est donnée par une loi de puissance avec un indice $\mu < 2$, on peut s'attendre à ce que le herfindahl H soit supérieur à $1/N$.

Dans ce cas, les variables $x_i = \left(\frac{s_i(t)}{S(t)}\right)^2$ sont des lois de puissance distribuées avec un exposant $\mu/2$ comme expliqué en détail dans la Sec. 2.3, et leur somme H peut s'écrire

$$H \propto N^{\frac{2}{\mu}-2} \xi_{L,\mu/2}, \tag{c.6}$$

avec $\xi_{L,\mu/2}$ une variable de Lévy de paramètre $\mu/2$.

Des éléments empiriques ont été avancés en faveur de cette explication, par exemple dans l'article original de Gabaix [120] mais aussi dans [131]. Cependant, une exploration empirique minutieuse permet de constater que d'importantes corrélations intersectorielles subsistent même à grand niveau de désagrégation, et que les *effets de réseau* sont en fait dominants [114, 2]. Ceci est en contradiction avec le point de vue proposé par l'hypothèse de "granularité", selon laquelle les entreprises subissent des chocs idiosyncrasiques indépendants. Ce constat explique donc la nécessité d'introduire des réseaux de production dans les modèles visant à expliquer la volatilité de l'économie.

c.2 LES MODÈLES DE RÉSEAUX DE PRODUCTION

Prenons une économie constituée de N entreprises. Alors une façon très simpliste de décrire la production π_i de l'entreprise i au moment t est d'écrire que

$$\pi_i = f_i\big(\{Q_{ij}, \quad j \in \mathcal{N}(i)\}\big), \tag{c.7}$$

avec $\mathcal{N}(i)$ l'ensemble des fournisseurs de l'entreprise i et Q_{ij} la quantité de biens produits par j que i a achetés dans le passé et qui sont utilisés au moment de la production. La fonction f_i est appelée une fonction de production et représente de manière phénoménologique la technologie de production de i. Cette équation correspond à une vision qu'on pourrait dire "métabolique" du rôle d'une entreprise, en tant qu'entité qui prend qui prend des choses en entrée et crache un produit avec une certaine efficacité.

Par exemple, on peut imaginer que i est un constructeur automobile et que $\mathcal{N}(i)$ est l'ensemble des entreprises j et des ménages qui vendent la main-d'œuvre et les matériaux nécessaires à la production d'une voiture, c'est-à-dire les pneus, les pare-brise, les écrous, les boulons, etc., mais aussi des heures de travail. La fonction de production f_i nous indique alors combien de voitures peuvent être fabriquées avec des quantités Q_{ij} en entrée.

L'idée derrière tout cela est de constater que les entreprises sont entremêlées, la baisse de production d'une entreprise donnée peut s'amplifier

tout le long de la chaîne d'approvisionnement et perturber la production de ses clients. Inversement, une baisse de la demande se propagerait dans la direction opposée. Ce scénario a été proposé dans l'article de Long et Plosser [169], et a ensuite inspiré une grande partie des recherches que je décris par la suite. Il convient de noter que l'analyse des entrées-sorties, *input-output analysis* en anglais, a historiquement été initiée par l'économiste russe Leontief [165], qui s'est inspiré des idées du Tableau économique de Quesnay [223].

Plus récemment, Acemoğlu *et al.* ont repris le cadre de Long et Plosser pour expliquer des fluctuations économiques anormales, dans un article qui reprend tout de même plusieurs idées de la "granularité" [4]. Dans leur modèle, la fonction de production est du type Cobb-Douglas, et est donc donnée par l'équation

$$\pi_i = z_i^{\alpha} \ell_i^{\alpha} \prod_j Q_{ij}^{(1-\alpha)w_{ij}}, \tag{c.8}$$

où z_i est un "choc de productivité", ce qui selon leur interprétation signifie une augmentation ou diminution momentanée de l'efficacité de l'entreprise. Pour les autres variables, ℓ représente le travail, α est la part du travail dans la production de i, Q_{ij} est la quantité de biens produits par j intervenant dans la production de i et w_{ij} est un coefficient de pondération définissant le tableau d'entrées-sorties, ou *input-output table* en anglais. Ce tableau st également représenté par une matrice $\mathbf{W}$ telle que $(\mathbf{W})_{ij} = w_{ij}$, dont on suppose que $\sum_j w_{ij} = 1$. Une explication plus détaillée de Cobb-Douglas et de la famille plus générale des fonctions de production de CES est fournie en Annexe D. Un aspect crucial de cette fonction de production, comme souligné dans l'appendice, est que les biens peuvent être substitués entre eux : si on baisse une certaine quantité comme $Q_{ij} \to f Q_{ij}$, avec $f < 1$, alors le niveau de production π_i peut être maintenu à condition de multiplier une autre quantité en entrée Q_{ik} par $f^{-\frac{w_{ik}}{w_{ij}}}$.

Ils supposent également qu'il y a un ménage représentatif qui consomme une partie de la production des entreprises en maximisant une fonction d'utilité sous contrainte budgétaire. En écrivant des équations d'équilibre où les entreprises maximisent leurs profits, les auteurs montrent que le PIB S de l'économie est donné par

$$\ln(S) = \sum_i v_i \varepsilon_i, \tag{c.9}$$

avec $\varepsilon_i = \ln(z_i)$ et v_i le *vecteur d'influence/"influence vector"* qui mesure la part des ventes de l'entreprise i dans l'économie à l'équilibre, à savoir

$$v_i = \frac{p_i \pi_i}{\sum_j p_j \pi_j} = \frac{\alpha}{N} \sum_j (\mathbf{L})_{ij}, \tag{c.10}$$

où $\mathbf{L} = (\mathbf{1} - (1-\alpha)\mathbf{W})^{-1}$ est appelé l'*inverse de Leontief*.

Il existe alors un parallèle évident entre les équations (c.4) et (c.9) si l'on compare le logarithme des chocs de productivité ε_i avec $\ln(1 + m + \sigma\eta_i)$.

Dans la Réf. [4] les auteurs voient le tableau **W** comme la matrice de connectivité pondérée d'un réseau, et ils prouvent ensuite que si le réseau est très déséquilibré, c'est-à-dire avec une distribution des degrés des nœuds du réseau en loi de puissance de puissance ou dans le cas d'une économie "étoilée" où une entreprise approvisionne toutes les autres, alors la distribution de v_i a des queues lourdes. Ainsi, le cadre de la granularité s'applique et on explique les fluctuations anomales *seulement si* le réseau satisfait aux conditions nécessaires pour que la distribution des v_i mène à des phénomènes de concentration. Il est donc, à mon avis, nécessaire d'aller au delà de ce cadre et de proposer une explication plus générale.

c.3 RÉSEAUX DE PRODUCTION MARGINALEMENT STABLES

Dans ce qui précède, j'ai présenté deux explications à ce que Ben Bernanke a appelé en anglais le "*small shocks, large business cycle puzzle*" . Une piste différente consisterait à considérer que les nombreux chocs affectant l'économie sont en fait de nature endogène, et qu'en fait l'économie s'auto-organise vers un état instable et turbulent. Cette idée a d'abord, à ma connaissance, été mentionnée en 1948 par Hawkins [142], et reprise par Bak, Chen, Scheinkman et Woodford [34, 231], avec ces derniers s'appuyant sur les idées de la criticalité auto-organisée (dont l'acronyme en anglais est SOC pour *self-organized criticality*) examinée en détail dans la section 3.4.4. Comme un tas de sable qui continue à croître vers une position instable, l'économie, selon eux, tend vers des états où elle est très susceptible d'amplifier de chocs.

Le modèle suivant présente des idées similaires, et il s'agit d'une contribution originale que j'ai publiée avec Jean-Philippe Bouchaud dans [191] et qui étend les modèles d'économie de réseau. La différence majeure avec les travaux récents se trouve dans l'utilisation de fonctions de production plus générales que celles de Cobb-Douglas, avec pour caractéristique principale que les différentes entrées de la production d'une entreprise ne peuvent pas être facilement substitués. En conséquence, l'existence d'un équilibre concurrentiel, comme celui dont il est question chez Acemoğlu et ses collaborateurs [4], n'est pas toujours garantie, mais dépend d'une contrainte sur le spectre des matrices qui décrivent l'interaction entre les entreprises. C'est ce qui est appelé la condition de Hawkins-Simon [142, 143, 112, 217, 220].

c.3.1 *Description du modèle*

Je considère dans cette partie un réseau de N entreprises, modélisé comme un graphe dirigé où les nœuds représentent les entreprises et où un lien $j \to i$ existe si i a besoin de biens produits par j dans sa production. Le nœud $i = 0$ représente les ménages, qui fournissent leur travail tout en consommant une partie de la production des entreprises. Le lien $j \to i$ porte un "poids stœchiométrique" J_{ij}, qui mesure le nombre de biens j

nécessaires pour fabriquer une unité de i. L'ensemble des fournisseurs de i est donc donné par $\{j/J_{ij} \neq 0\}$, tandis que l'ensemble de ses clients est $\{j/J_{ji} \neq 0\}$.

La production de i, π_i, est donnée en toute généralité par une fonction dite CES (*Constant Elasticity of Substitution*, fonction à élasticité de substitution constante), mais il est plus commode d'utiliser une fonction de Leontief, en sachant que les arguments développés ici restent inchangés si on considère toute autre fonction de la famille CES, à part la fonction de Cobb-Douglas. La fonction de production de Leontief est

$$\pi_i = z_i \min_j \left(\frac{Q_{ij}}{J_{ij}} \right). \tag{c.11}$$

Dans ce cas aucun bien n'est substituable : si on imagine que i produit des voitures et que j produit des pneus, alors J_{ij} vaut 4, et si on n'a que 4 pneus on ne pourra construire qu'une voiture, même si on a des milliers de pare-brises. Ce cas Leontief est donc un modèle ou il est difficile de se passer de certains biens ; les entreprises choisissent leurs fournisseurs avec parcimonie et ne peuvent en changer facilement. C'est un cadre réaliste, qui correspond à ce qui s'est passé au Japon après le tsunami et le désastre de Fukushima en 2011, mais aussi à ce qui a eu lieu passé récemment avec les matériaux nécessaires à la production de masques de protection respiratoire et aux réactifs pour les tests épidémiologiques lors de la crise du Covid-19, qui est encore en cours au moment où j'écris ces lignes.

On peut alors montrer que les prix d'équilibre satisfont à une équation

$$\mathscr{P}_i = 0 \longrightarrow z_i p_i - \sum_{j \neq 0} J_{ij} p_j = V_i \quad (> 0), \tag{c.12}$$

où V_i est déterminé par les caractéristiques des ménages. De même, on peut écrire la production d'une entreprise comme $\pi_i := z_i \gamma_i$, où les γ_i sont des *niveaux de production*. À l'équilibre, la production d'une entreprise est intégralement achetée par ses clients, et donc les niveaux de production équations du type

$$\pi_i = \sum_{j \neq 0} Q_{ji} + C_i \longrightarrow z_i \gamma_i - \sum_{j \neq 0} J_{ji} \gamma_j = C_i \quad (> 0), \tag{c.13}$$

où les C_i sont déterminés par la consommation des ménages.

Dans les deux cas, on peut écrire des équations matricielles,

$$\begin{aligned} \mathbf{M} \, |p\rangle &= |V\rangle \\ \mathbf{M}^t \, |\gamma\rangle &= |C\rangle \, , \end{aligned} \tag{c.14}$$

où la matrice $\mathbf{M}$ est telle que $(\mathbf{M})_{ij} = z_i \delta_{ij} - J_{ij}$.

Ces deux équations doivent avoir des solutions de prix et de niveaux de production *positives*. La forme de $\mathbf{M}$ est particulière, puisqu'elle n'a que des termes positifs sa diagonale et des termes négatifs ou nuls ailleurs, et

il s'agit alors d'une Z-matrice [112]. La condition de Hawkins et Simon veut alors que ce soit le cas si et seulement si la matrice $\mathbf{M}$ a un spectre dont la partie réelle est positive, c'est-à-dire qu'elle doit appartenir à un sous-ensemble des Z-matrices qu'on appelle les M-matrices.

c.3.2 *Une illustration simple*

Pour se faire une intuition de cette condition, considérons un réseau aléatoire régulier. Ainsi, chaque lien J_{ij} est égal à J avec une probabilité r et à 0 avec une probabilité $1-r$, et on prend également toutes les entreprises N comme ayant même productivité z.

Le spectre de $\mathbf{M}$ dans ce cas est bien connu lorsque $N \gg 1$ et $r \sim \mathcal{O}(1)$. Il se compose d'une valeur propre isolée $\lambda_{\min} = z - JrN$ et d'une "mer" de valeurs propres complexes uniformément distribuées dans un disque de rayon $J\sqrt{r(1-r)N}$ et centré en z (voir e. g. [132, 260], et comparer avec le cas symétrique discuté dans la section C.1.3).

La condition de stabilité se lit alors $z > JrN$ et implique que la productivité doit être suffisamment importante pour que l'économie fonctionne. Le vecteur propre le plus instable, correspondant à la valeur propre $\lambda_{\min}$, est un vecteur uniforme de coordonnées $(1/\sqrt{N}, \ldots, 1/\sqrt{N})$. Comme on le verra plus loin, cela correspond à un cas où les crises sont systémiques. Le même résultat qualitatif est valable lorsque les productivités sont faiblement hétérogènes, soit $z_i = z(1+\varepsilon_i)$ avec $\varepsilon_i \ll 1$ (bien que $\lambda_{\min}$ soit légèrement décalé vers la gauche d'une quantité $O(z^2\varepsilon^2/JN)$).

On retrouve un cas plus intéressant, mais aussi plus complexe, lorsque le nombre moyen de fournisseurs de chaque entreprise $c = rN$ est de l'ordre de l'unité (c'est-à-dire quand $r = \mathcal{O}(N^{-1})$). Dans le cas du réseau régulier aléatoire (RRN, pour *random regular network* en anglais) où chaque entreprise a exactement c fournisseurs (et c clients) choisis au hasard parmi les $N-1$ autres entreprises, on sait que le spectre de $\mathbf{M}$ est à nouveau constitué d'une valeur propre isolée $\lambda_{\min} = z - Jc$ et d'une "mer" de valeurs propres complexes distribuées dans un disque de rayon $J\sqrt{c}$ centré en z. [1] Le lecteur désirant avoir une intuition plus précise dans le cas où la matrice est symétrique, soit $J_{ij} = J_{ji}$, et dans le cas où toutes les entreprises ont le même nombre de clients/fournisseurs, peut consulter l'annexe C.

Lorsqu'on considère des économies avec un certain degré d'hétérogénéité, soit topologique (en laissant le nombre de fournisseurs ou de clients varier), soit parce que les couplages J et les productivités z fluctuent, il n'est plus possible de trouver des résultats exacts. En revanche, pour une matrice $\mathbf{M}$ symétrique, les résultats exacts restent encore rares mais un travail énorme a été fait en physique, mathématiques et informatique pour caractériser les valeurs propres et les vecteurs propres de ces matrices [272, 46, 108, 227, 118, 161, 43, 198, 48, 5]. En effet, ces familles de matrices aléatoires symétriques apparaissent dans de nombreuses situations physiques, comme

1 Dans ce cas, cependant, la densité des valeurs propres complexes n'est pas uniforme mais est donnée par $\rho(\lambda) \propto (c^2 - |\lambda|^2)^{-2}$ pour $|\lambda| < \sqrt{c}$ [183].

l'étude du spectre de vibration de solides amorphes ou le spectre d'énergie de systèmes quantiques avec impuretés. Elles apparaissent également en théorie des graphes et en informatique. Réussir à estimer la valeur propre extrême est un problème particulièrement important, car il apparaît dans de nombreux problèmes différents (par exemple la propagation d'épidémies ou de rumeurs [81], voir aussi [253] — ou la propagation de crises comme dans cette thèse) ; par ailleurs, le vecteur propre associé est lié au concept de centralité en théorie des réseaux [199] ; voir aussi [80] et les références qui s'y trouvent.

Dans la suite, Δ^2 donne un ordre de grandeur de la variance des fluctuations (de connectivité, de productivité, etc.). On peut alors résumer les résultats des dernières années dans le le scénario suivant, amplement décrit dans [48] et visibles dans la Figure c.2 dans le cas d'un réseau régulier :

- Si $\Delta = 0$, toutes les valeurs propres sauf une ont une partie réelle confinée dans un certain intervalle $\mathscr{I}$ ($= [z - \sqrt{c}, z + \sqrt{c}]$ dans l'exemple RRN), alors que la valeur propre isolée est située à gauche de cet intervalle, à une distance non nulle g (comme *gap* en anglais) de son bord.

- Au fur et à mesure que Δ augmente, l'intervalle $\mathscr{I}$ s'élargit et ses bords deviennent diffus, tandis que la valeur propre isolée se rapproche de plus en plus du bord inférieur de $\mathscr{I}$, comme indiqué dans la Figure c.2.

- Au-delà d'une certaine valeur critique Δ_c, la valeur propre isolée est absorbée par $\mathscr{I}$ et disparaît (c'est ce qu'on appelle, dans un autre contexte, la transition Baik-Ben Arous-Péché (BBP) [30]). Le mécanisme derrière cette transition peut se comprendre par le mouvement brownien de Dyson, décrit en Sec. C.2 et surtout dans la Figure C.5.

En outre, dès que Δ est non nul, l'intervalle $\mathscr{I}$ doit être subdivisé en 3 intervalles $\mathscr{I}_-, \mathscr{I}_0, \mathscr{I}_+$ (avec $\mathscr{I}_0$ éventuellement vide, voir Fig. c.2), qui correspondent à des structures des vecteurs propres très différentes. Dans la partie centrale $\mathscr{I}_0$, les vecteurs propres sont *étendus*, ou *délocalisés*, alors que dans les parties extrêmes $\mathscr{I}_-, \mathscr{I}_+$, les vecteurs propres sont *localisés*. De manière générale, on dit qu'un vecteur est "localisé" quand la plupart de ses composantes sont concentrées sur quelques nœuds (entreprises) seulement, tandis que "délocalisé" signifie que les valeurs des composantes sont bien réparties sur tous les nœuds.

Plus précisément, si on note $v_1, v_2, \ldots, v_N$ les composantes d'un vecteur normalisé $|V\rangle$, la nature localisée/délocalisée de $|V\rangle$ se comprend par herfindahl H (appelé *inverse participation ratio* (IPR) en physique de la matière condensée) :

$$H(|V\rangle) = \sum_i^N |v_i|^4. \tag{c.15}$$

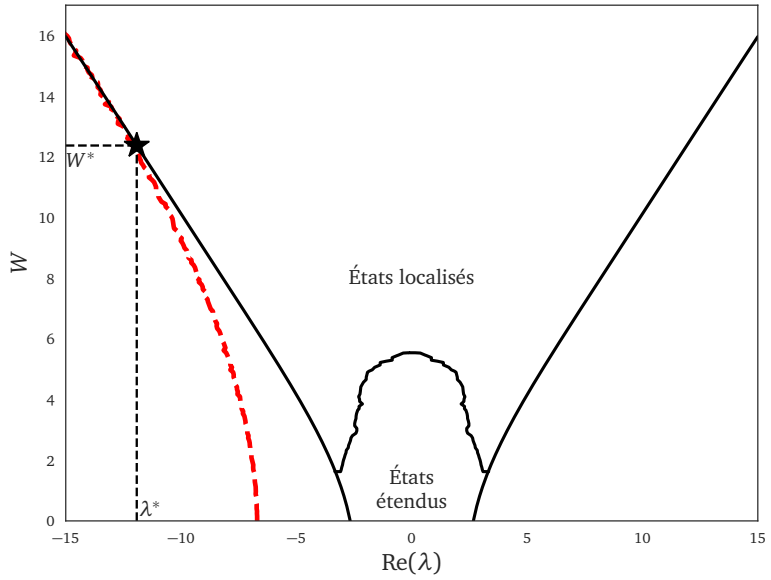


FIGURE c.2 : Résultats numériques décrivant la structure des états propres d'un réseau régulier aléatoire dirigé (RRN) avec $N = 2000$ entreprises ayant chacune $c = 7$ fournisseurs et clients et avec $J = 1$. Les productivités z_i sont uniformément distribuées dans un intervalle $[z - W/2, z + W/2]$, et l'axe des abscisses est centré autour de z. Les états propres contenus dans la zone du milieu se localisent au fur et à mesure que W augmente. La courbe rouge pointillée à gauche correspond à la valeur propre isolée qui finit par être absorbée au point marqué par une étoile noire, correspondant ici à $W^* \approx 12.4$ et $\lambda^* \approx -12$. La frontière entre les états étendus et localisés est définie ici par $H = 5/N$. Comparer avec la Figure 1. dans [48], pour le cas des RRN non dirigés.

Comme pour le herfindahl "normal", un vecteur propre localisé est tel que $H(|V\rangle) = \mathcal{O}(1)$ dans la limite $N \to \infty$ alors qu'un vecteur propre délocalisé a un herfindahl $H(|V\rangle) = \mathcal{O}(N^{-1})$.

c.3.3 *Propagation et amplification de crises*

Dans cette partie, je tâcherai de montrer l'importance de la distinction qui a été faite entre états localisés et délocalisés. En raison de la structure de $\mathbf{M}$, le théorème de Perron-Frobenius nous dit que la valeur propre la plus à gauche $\lambda_{\min}$ est réelle et que le vecteur propre associé a des composantes réelles positives $u_i > 0$. Comme indiqué précédemment, $\lambda_{\min}$ doit être positif pour que $\mathbf{M}$ soit une M-matrice, c'est-à-dire pour que tous les prix et tous les niveaux de production soient positifs à l'équilibre. Quand $\lambda_{\min} \to 0$, l'économie devient de plus en plus fragile et susceptible aux chocs extérieurs.

Considérons donc le cas où la productivité z_i de certaines entreprises diminue de $-\varepsilon\Delta_i < 0$, et/ou que certains des poids J_{ij} augmentent de $\varepsilon\Delta_{ij} > 0$. En utilisant la théorie des perturbations au premier ordre en ε, on constate que la valeur propre la plus à gauche est décalée de :

$$\lambda_{\min} \longrightarrow \lambda_{\min} - \varepsilon \left[\sum_i \Delta_i u_i^2 + \sum_{i \neq j} \Delta_{ij} u_i u_j \right]. \tag{c.16}$$

Comme les deux termes sont négatifs, cette formule montre qu'à mesure que le système devient marginalement stable, toute diminution locale de la productivité ou augmentation des poids J_{ij} peut faire basculer le système vers la région instable. Un certain nombre de prix/quantités deviennent alors négatifs. Intuitivement, on comprend déjà que la physionomie de ces "crises" dépend de la nature localisée/délocalisée du vecteur propre $|U\rangle$ correspondant à $\lambda_{\min}$.

La terme de second ordre dans l'équation précédente est d'ordre ε^2/g, où g est l'écart, ou *gap*, entre $\lambda_{\min}$ et la valeur propre suivante de $\mathbf{M}$; la théorie des perturbations au premier ordre n'est donc valable qu'à condition que $\varepsilon \ll g$. Il faut ensuite distinguer deux cas, selon la valeur des hétérogénéités :

- Si $\Delta < \Delta_c$, la valeur propre la plus à gauche est isolée, dans ce cas $g > 0$ même lorsque $N \to \infty$. Le résultat du premier ordre reste valable pour ε suffisamment petit. De plus, le vecteur propre associé $|U\rangle$ est délocalisé. On en déduit qu'un choc localisé — disons sur la seule firme ℓ — diminue $\lambda_{\min}$ de $\sim -\delta z_\ell / N$. Le système est donc instable pour $\delta z_\ell > N\lambda_{\min}$, mais pour que cette condition soit compatible avec $\delta z_\ell \ll g$, il faut aussi que $N\lambda_{\min} \ll g$. Ainsi, le choc se propage sur l'ensemble du système, en raison de la nature délocalisée de $|U\rangle$. Dans le cas d'un petit choc de productivité $\delta z_i = \delta z, \forall i$, la déstabilisation se produit dès que $\delta z > \lambda_{\min}$.

- Pour $\Delta > \Delta_c$, la valeur propre la plus à gauche se trouve au bord de l'intervalle $\mathscr{I}_-$, de sorte que l'écart $g(N)$ devient nul lorsque $N \to \infty$. En outre, le vecteur propre associé $|U\rangle$ est maintenant localisé, généralement centré autour d'entreprises à productivité particulièrement faible ou à forte connectivité (appelées régions de Lifschitz en physique de la matière condensée [167, 262, 48]). Or dans ce cas la théorie des perturbations de premier ordre s'effondre quand $\varepsilon \sim g(N)$, et il faut donc avoir recours à des simulations numériques pour caractériser les crises associés.

Afin de comprendre intuitivement cette réponse aux perturbations proche du point critique $\lambda_{\min} \approx 0$, regardons le cas simple où $\forall i$, $z_i = z$. On peut développer $\mathbf{M}^{-1}$ dans la région stable comme

$$\mathbf{M}^{-1} = \frac{1}{z} \sum_{k=0}^{\infty} \left(\frac{\mathbf{J}}{z} \right)^k \tag{c.17}$$

avec $(\mathbf{J})_{ij} = J_{ij}$, puisque la condition de stabilité implique que le rayon spectral de $\mathbf{J}$ est inférieur à z et justifie alors ce développement en série. Le terme $(\mathbf{J}^k)_{ij}$ n'est autre que la somme de tous les chemins de longueur k reliant la firme j à la firme i. La stabilité marginale correspond au cas où cette somme devient de plus en plus divergente, et donc quand des chemins de longueur arbitraire contribuent à $(\mathbf{M}^{-1})_{ij}$. Cette interprétation est également valable dans le cas de z_is hétérogènes. [2] L'instabilité est donc liée à une situation où les chocs peuvent se propager sur des chemins de longueur arbitraire dans le réseau. Ce cas de figure est lié aux transitions de phase du second ordre en physique, où les corrélations s'étendent sur des distances macroscopiques de taille arbitraire et où la réponse à de petites perturbations est aussi divergente, voir par exemple [234] et la discussion détaillée de la section 3.4.5.

c.3.4 *Résultats numériques*

J'étudie ici un modèle très simple de réseau régulier aléatoire, où chaque entreprise a exactement c fournisseurs et c clients, chacun choisi au hasard parmi les $N-1$ autres entreprises. Chaque entreprise a une productivité aléatoire tirée uniformément dans l'intervalle $[z-W/2, z+W/2]$, de sorte que z est la productivité moyenne et $\Delta = W/2\sqrt{3}$. Les coefficients stœchiométriques J_{ij} sont tous égaux à $J = 1$. En outre, je fixe $V_i = 1$ pour simplifier et je prends la consommation des ménages $C_i = 1/p_i$, obtenue à partir d'une fonction d'utilité logarithmique avec des préférences identiques pour tous les produits. Je me place ensuite sur le point critique, en choisissant z et W tels que $\lambda_{\min} = \varepsilon = 10^{-8}$.

Deux quantités sont alors particulièrement intéressantes ; la distribution des tailles des entreprises, définie comme leurs ventes $\mathscr{S}_i = z_i \gamma_i p_i$ et la distribution de taille des "avalanches" de faillites suite à un choc de productivité exogène. On peut remarquer qu'en présence d'un désordre W suffisant les deux distributions présentent des queues qu'on peut apparenter à des lois de puissance.

En outre, il est possible de faire une typologie des crises selon la valeur prise par W et la nature – localisée ou délocalisée – des vecteurs propres proches du bord du spectre. Ainsi, un vecteur instable délocalisé et isolé de la partie continue du spectre correspond à une crise qui frappe toute l'économie, alors que s'il est localisé la crise ne touchera qu'un sous-secteur de l'économie. Pour une discussion détaillée des résultats numériques de ce modèle, voir la partie 5.3.

2 En effet, on peut toujours écrire $\mathbf{M} = z_{\max}\mathbf{1} - \mathbf{B}$ avec $\mathbf{B}$ une matrice à composantes positives, et on peut encore développer en série comme $\mathbf{M}^{-1} = \frac{1}{z_{\max}} \sum_{k=0}^{\infty} \left(\frac{\mathbf{B}}{z_{\max}}\right)^k$.

c.3.5 *Stabilité marginale*

Il reste encore à justifier pourquoi une économie pourrait être dans un état "critique", avec une valeur propre minimale $\lambda_{\min} \approx 0$. Je vais donc expliquer pourquoi une économie pourrait de façon générique — comme beaucoup d'autres systèmes complexes, voir e. g. [32, 103, 84, 195, 47] — s'"auto-organiser" pour être à la limite de la stabilité, dans une situation marginalement stable où $\lambda_{\min} \to 0$. Plusieurs mécanismes pourraient être la raison derrière cela. C'est le cas de l'introduction dans l'économie de nouvelles entreprises, des baisses de productivité, de la complexification des liens inter-firmes, de l'optimisation de porte-feuilles de fournisseurs et de l'exigence de certaines entreprises de réaliser des marges. Tout cela peut se comprendre aisément par l'analyse spectrale, via la théorie des matrices aléatoires, de la matrice $\mathbf{M}$.

c.3.6 *Vers un modèle dynamique*

De nombreuses pistes restent à explorer dans ce modèle. Un indice prometteur revient à doter le modèle d'une dynamique réaliste, en partie inspirée du modèle d'instabilités de la référence [52], qui inclurait des frictions, le caractère myope du comportement des agents, le recâblage, etc. En plus de rendre le modèle plus réaliste, avoir un modèle dynamique est une condition préalable au calibrage sur des données empiriques.

Ceci constituera la base d'un prochain article [101], réalisé en collaboration avec Théo Dessertaine, Michael Benzaquen et Jean-Philippe Bouchaud. L'idée principale est que les entreprises produisent chacune un bien d'une certaine façon, capturée de manière phénoménologique par une fonction de production. Les firmes doivent ensuite décider d'une stratégie pour fixer leur prix, leur niveaux de production et les achats à faire à leur fournisseur, et le modèle peut ensuite être étudié avec des outils issus de la théorie des systèmes dynamiques.

c.4 CONCLUSION ET PERSPECTIVES

Les modèles présentés ci-dessus fournissent des indices théoriques sur les mécanismes possibles de propagation et d'amplification de fluctuations dans l'économie. Selon l'argument dit de la granularité, l'hétérogénéité dans la taille des entreprises d'un pays est responsable d'une part importante des fluctuations macroéconomiques observées, et la discussion qui lui a succédé sur les modèles en réseau a montré la nécessité de prendre en compte les interconnexions entre les différentes entreprises.

Il s'agit donc de modèles qui tentent d'aller au-delà de l'idée de chocs indépendants exogènes qui touchent les entreprises. D'une certaine manière, cela ressemble beaucoup aux progrès réalisés après le modèle initial de Louis Bachelier pour les fluctuations des cours boursiers sous forme de

mouvement brownien géométrique [28], où les prix p_t sont modélisés comme des mouvements browniens géométriques

$$\ln(p_{t+1}) - \ln(p_t) = m + \sigma\eta_t := r_t, \tag{c.18}$$

où r_t est appelé le *return* en finance et encode les fluctuations aléatoires du prix. Il se mesure en pour-cent par unité de temps, et a une moyenne de m, et une volatilité σ^2. Le dernier terme η_t est une variable aléatoire qui est prise comme gaussienne dans le modèle de Bachelier.

Malgré son utilisation généralisée en mathématiques financières, et malgré son ubiquité dans beaucoup de salles de cours, ce modèle s'avère empiriquement trop simple et ne peut expliquer une longue série de faits stylisés présents dans les données empiriques. Le lecteur pourra trouver une liste de ces faits stylisés dans la Réf. [93], mais certaines des caractéristiques importants que l'on trouve dans les séries temporelles de *returns* sont leur distribution en queue lourde (ce qui signifie que la variable η ne doit pas être prise comme gaussienne) et leur intermittence (ce qui signifie qu'il y a de temps à autre des rafales turbulentes dans les séries temporelles).

Des travaux prometteurs ont ainsi été faits dans le but d'expliquer ces faits statistiques et de comprendre comment ils peuvent émerger des actions microscopiques des agents qui interagissent sur les marchés financiers [66, 67]. Une piste intéressante a été proposée récemment dans les articles [152, 15], où les auteurs proposent de modéliser les prix des actifs financiers comme évoluant via une dynamique où ils interagissent de façon non-linéaire et où l'on ajoute aussi du bruit stochastique. Leur méthode permet d'obtenir des résultats analytiques et numériques, montrant ainsi que l'ajout d'un couplage entre différents actifs avec des boucles de rétroaction non linéaires permet de récupérer des caractéristiques comme la volatilité intermittente, et de reproduire certaines des statistiques des *returns* r_t définis dans l'Éq. (c.18). Des résultats similaires ont été obtenus pour l'étude des cycles économiques dans la Réf. [207].

Au vu de ces succès scientifiques en finance, et afin de mieux comprendre les grandes fluctuations macroéconomiques, il est clair qu'il serait souhaitable de trouver des mécanismes similaires affectant les variables dites *fondamentales* décrivant une entreprise, c'est-à-dire celles liées à sa production, ses ventes ou d'autres variables non liées au marché. En finance cependant, beaucoup de ces résultats reposent sur une base bien documentée de faits stylisés qui deviennent des contraintes, puisque tout modèle réaliste devrait les reproduire. Ainsi des chercheurs ont étudié la statistique des taux de croissance qui décrivent les dynamiques de croissance des entreprises (voir les références [251, 164, 58] et autres qui sont fournies dans le Chapitre 6). Ainsi le but du Chapitre 6, et de la partie suivante qui le résume, est d'explorer ces modèles et de comprendre la nature des faits stylisés qui décrivent la croissance des entreprises.

d

ÉTUDE EMPIRIQUES DES TAUX DE CROISSANCE D'ENTREPRISES

Un fait stylisé bien établi dans la description des entreprises, comme dit dans la partie précédente, est que leur "taille", définie par leurs ventes ou de leur capitalisation boursière par exemple, est bien décrite par une loi de Zipf, à savoir une loi de puissance avec un exposant $\mu \approx 1$. Cela a d'abord été remarqué en 1931 par Robert Gibrat [128], et Simon [244] a ensuite remarqué que ce n'est ce à quoi on s'attendrait, par exemple, s'il existait une "taille optimale" pour une entreprise, qui pourrait être calculée en optimisant des fonctions d'utilité ou de profits sous certains contraintes.

À la place d'un tel modèle d'optimisation rationnelle, Gibrat a proposé un "modèle de croissance stochastique" et multiplicative, basé sur de la croissance multiplicative qui peut conduire, comme je montre dans le Chapitre 3, à des lois de puissance.

C'est ce qu'on appelle la "loi de Gibrat", dont je donne maintenant une définition plus précise. Considérons une certaine quantité x décrivant quantitativement une entreprise ; il peut s'agir de sa capitalisation boursière ou de son chiffre d'affaires net, [1] et on note alors $x_i(t)$ sa valeur pour l'entreprise i à l'instant t. :e modèle de croissance multiplicative de Gibrat préconise alors une dynamique donnée par l'équation suivante :

$$x_i(t+\Delta t) = \Big(1 + g^i_{x,\Delta t}(t)\Big) x_i(t) \tag{d.1}$$

où $g^i_{x,\Delta t}$ est défini comme le taux de croissance (en pourcentage) de la quantité x_i entre t et $t+\Delta t$. Lorsque ce taux g est faible par rapport à 1, ce qui est le cas de façon générale, on peut l'identifier au taux de croissance *logarithmique*,

$$g^i_{x,\Delta t}(t) = \log\left(\frac{x_i(t+\Delta t)}{x_i(t)}\right), \tag{d.2}$$

qui constitue la définition habituelle du taux de croissance que j'adopte par la suite.

La version forte de la loi de Gibrat dit que la statistique de g est indépendante de l'entreprise, et notamment de sa taille. Un nombre considérable de preuves contredit néanmoins cette hypothèse (voir [12] et les références qui y figurent). Ainsi, on constate par exemple que l'écart-type de g dépend de x, comme

$$\text{std.}(g|x) := \sigma(x) \propto x^{-\beta}. \tag{d.3}$$

.

1 La quantité précise que l'on utilise, qu'il s'agisse d'un flux ou d'un stock, ne change pas l'analyse. Voir par exemple la Réf. [20].

Beaucoup de chercheurs ont également travaillé sur la distribution des taux de croissance, $p(g)$ [251, 12], et un grand nombre de travaux parlent de distributions de Laplace, de Subbotin ou autres distributions similaires qui ont toutes pour caractéristique communes d'avoir une forme en "tente " [57, 58, 56, 55]. En ce qui concerne les queues de la distribution, d'aucuns disent aussi qu'elles sont bien décrites par des lois de puissances – même par des lois de Cauchy [274], qui ne sont pas dans le bassin gaussien du théorème central limite – ou des exponentielles étirées.

Pour ma thèse, j'ai essayé de reprendre ce genre de travaux, en m'inspirant de ce qui a été fait dans l'analyse de séries de données financières (par exemple dans [66, 67]). Le lecteur comprendra rapidement qu'il y a malgré tout une distinction cruciale à faire entre les données financières et les données "fondamentales" d'entreprises. Les premières sont abondantes, et on dispose généralement de très longues séries chronologiques pour chaque actif financier. En revanche, les données au niveau de l'entreprise sont souvent constituées de séries chronologiques relativement courtes (quelques dizaines de points de données au maximum) qu'il faut manipuler avec précaution. Il faut donc traiter correctement ces dernières, et la prise en compte de la nature hétérogène de la volatilité est, je pense, ma principale contribution à ce domaine. Les travaux que je décris en détail dans le Chapitre 6 n'ont pas encore été publiés, et ont été réalisés en collaboration avec Angelo Secchi et Jean-Philippe Bouchaud.

d.1 DESCRIPTION DES DONNÉES

Le travail que j'ai fait a été une étude comparative entre la dynamique des variables fondamentales et les variables financières qui décrivent les entreprises ; j'ai donc utilisé quatre sources de données différentes. La première est la base de données CompuStat/CRSP, où je n'ai conservé que les entreprises dont les actions sont cotées sur le NASDAQ et le NYSE ; j'ai également converti toutes les valeurs en dollars américains à l'aide des données du FOREX ; l'intervalle de temps considéré dans mon étude va de 1962 à 2013 et correspond à un échantillon de 20140 entreprises, avec des données ayant une fréquence trimestrielle. Cet fréquence relativement élevée m'a conduit à utiliser ce jeu de données comme le principal pour mon étude.

Le deuxième jeu de données est le jeu de données FICUS de l'INSEE, qui va de 1994 à 2007 et qui contient toutes les "entreprises" françaises enregistrées pendant cette période, à l'exception de celles soumises à des contraintes de confidentialité en raison de leur secteur d'activité. La définition d'une entreprise dans cet ensemble de données est très large et correspond au système SIREN. Il s'agit d'un code unique à 9 chiffres utilisé pour identifier les entreprises françaises. Dans ce système, des unités différentes ayant des numéros SIREN distincts peuvent en fait faire partie d'une seule et même entreprise : par exemple, Total SA, Total Holdings SAS SNEA et Total Marketing correspondent toutes à des identifiants SIREN distincts

et sont donc considérées comme des unités différentes dans cet ensemble de données, bien qu'elles fassent toutes partie de la même entreprise.

En partant d'un groupe initial d'environ 6 millions d'entreprises, j'ai sélectionné uniquement celles qui étaient présentes tous les ans entre 1994 et 2007 et qui ont déclaré au moins un employé pendant cette période. Au total, ce sous-ensemble contient 374714 entreprises. Par ailleurs, la fréquence des données le long de cette durée est annuelle.

Je voudrais également attirer l'attention du lecteur sur le fait que l'ensemble de données de l'INSEE est constitué d'entrées qui sont remplies par les entreprises elles-mêmes. Elles sont donc susceptibles de contenir des erreurs, ce qui a impliqué un travail important de nettoyage en amont. Ce travail m'a pris quelques mois, pour que je puisse être confiant qu'aucune entrée aberrante n'était présente dans les données. Il s'agissait la plupart du temps d'erreurs de saisie, où il était flagrant que le chiffre était faux car éloignée de plusieurs ordres de grandeur d'une valeur "typique", auquel cas l'entrée était rejetée, ou de dates saisies dans un format erroné, ce qui pouvait être corrigé. L'ensemble de mon analyse des données de l'INSEE a été fait grâce au Centre d'accès sécurisé aux données (CASD), qui permet d'avoir accès à des données confidentielles produites par des entités publiques françaises. La référence de mon projet est REGSTAT.

Pour le troisième et quatrième jeux de données j'ai étudié les données *FactSet Supply Chain Relationships* et *Reuters Fundamentals*. Le premier est une liste de données relationnelle, indiquant si les entreprises A et B ont un rapport client/fournisseur. Elle est construite à partir de sources publiques primaires telles que des bilans annuels SEC 10-K, des présentations aux investisseurs ou des communiqués de presse, et couvre environ 23000 sociétés cotées en bourse et compte plus de 325000 liens entre entreprises. Les données Reuters m'ont ensuite permis de faire correspondre les données relationnelles avec des variables fondamentales et financières des entreprises.

En utilisant FactSet, j'ai construit un graphe d'approvisionnement – *supply chain network* en anglais – comme cela est fait par ailleurs dans la section 5.4, en utilisant les liens inter-firmes datés de 2010 à 2019 puis en ne prenant que la plus grande composante faiblement connectée.[2] Le graphe obtenu est comporte 12257 entreprises que j'ai ensuite fait correspondre à la base de données Reuters, ce qui m'a permis d'obtenir des données échantillonnées avec une fréquence annuelle.

d.2 ÉTUDE EMPIRIQUE DES TAUX DE CROISSANCE

J'ai étudié le taux de croissance d'une quantité x en utilisant la définition donnée en équation (d.2). La plus grande fréquence d'échantillonnage correspond à celle présente dans les données CRSP/Compustat, et corres-

2 Un sous-graphe faiblement connecté est tel que deux nœuds quelconques peuvent être reliés par un chemin quelconque, par opposé à un graphe fortement connecté qui exige que ce chemin soit un chemin dirigé.

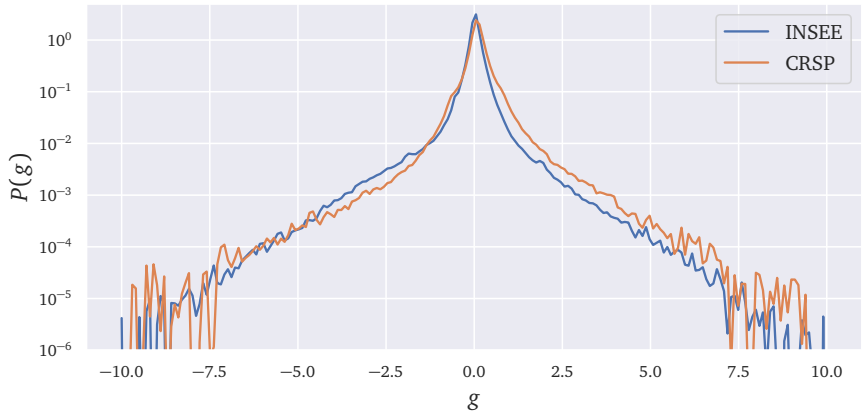


FIGURE d.1 : Distribution des taux de croissance annuels du chiffre d'affaires pour les firmes dans les bases de données INSEE et CRSP/COMPUSTAT. Le graphe correspond a une estimation des densités faite par convolution avec un noyau gaussien. Les deux ensembles de taux de croissance ont été centrés par leur moyenne globale et divisés par leur écart type, afin d'obtenir des variables centrées réduites.

pond à $\Delta t = 1$ trimestre. J'ai donc fixé l'unité de temps de mon étude à $\Delta t = 1$ trimestre, avec donc $g^i_{x,4}(t)$ le taux de croissance de $x_i(t)$ sur un an.

J'ai tracé la distribution des taux de croissance annuels du chiffre d'affaires, soit $g_{x,4}$ en prenant x le chiffre d'affaires, pour les données CRSP/-COMPUSTAT et INSEE sur la Figure d.1. Les deux distributions ont un "pic" rappelant une distribution de Laplace dans la région centrale, et des queues lourdes qui ressemblent à ceux de la distribution de Subbotin définie dans la Réf. [55], par exemple. Toutefois, ces caractéristiques sont toujours présentes, que l'on considère des taux de croissance sur un trimestre ou sur deux ans, ce qui n'est pas conforme à l'intuition qu'on pourrait avoir en prenant des variables de Laplace et en leur appliquant le théorème central limite, comme j'explique ci-dessous.

En effet, il est clair que le taux de croissance sur un an doit être statistiquement équivalent à la somme de quatre taux de croissance trimestriels, et donc de même loi que cette dernière. Cette observation n'est qu'une conséquence de l'égalité

$$g^i_{x,4}(t) = \sum_{k=0}^{3} g^i_{x,1}(t+k). \tag{d.4}$$

Or le pic "laplacien" qu'on observe dans les distributions comme celle en Figure d.1 ne survit pas à l'addition de variables aléatoires. En effet, comme je le montre dans l'Annexe A.2, on peut montrer que si g est la

somme centrée réduite de deux variables de Laplace, alors sa densité de probabilité est

$$p_{L,2}(g) = \frac{1}{2}(1+2|g|)e^{-2|g|} \underset{g\ll 1}{\approx} \frac{e^{-2g^2}}{2}, \tag{d.5}$$

qui ne présente pas de "pic" dans la région centrale. Il est possible de retrouver un résultat similaire pour la somme de variables de Laplace corrélées positivement ou négativement, comme je le montre dans l'Annexe A.2.2, avec le même résultat pour le pic.

Une région centrale comme celle qu'on voit en Fig. d.1 n'est donc pas compatible avec la structure donnée en équation (d.4), et elle est vraisemblablement la conséquence d'une mauvaise définition de g dans les études qui ont précédé la mienne.

On peut comprendre pourquoi une analyse des taux de croissance comme celle illustrée en Fig. d.1 échoue à l'examen précédent à l'aide de l'expérience de pensée suivante. Quand on a en tête une équation comme celle donnée en Eq. (d.2), on imagine en fait une sorte de simulation où on pourrait observer la firme i entre t et $t+\Delta t$ et regarder son évolution. On répéterait alors cette évolution plusieurs fois pour la *même* entreprise pour étudier la statistique de son taux de croissance. Si on prenait une période Δt plus grande, on devrait alors observer une région centrale qui devient de plus en plus gaussienne, comme on le voit dans l'étude de fluctuations de prix d'actifs financiers à plusieurs échelles de temps dans l'étude présentée dans la Réf. [218]. Je tâcherai maintenant de comprendre pourquoi les analyses qui ont précédé la mienne ont systématiquement indiqué des distributions en "forme de tente" pour les taux de croissance.

À mon sens, la raison derrière cela tient au fait que ces études ont mis, pour ainsi dire, toutes les entreprises dans le même panier. J'ai en effet déjà évoqué la dépendance de la volatilité des taux de croissance en la taille des entreprises. Pourquoi ne pas imaginer que cette volatilité est tout simplement hétérogène entre les entreprises ?

On peut en effet calculer la volatilité de chaque entreprise à l'aide des différentes observations de ses taux de croissance. Je définis alors la quantité

$$\sigma^i_{x,\Delta t} = \sqrt{\frac{\pi}{2}}\mathbb{E}_t\left\{\left|g^i_{x,\Delta t}(t) - \mathbb{E}_{t'}\left[\left|g^i_{x,\Delta t}(t')\right|\right]\right|\right\}, \tag{d.6}$$

où j'ai pris la moyenne de la valeur absolue de l'écart à la moyenne comme mesure de dispersion plutôt que l'écart type, ce dernier étant très bruité pour des séries temporelles peu longues comme celles avec lesquelles j'ai travaillé.

J'ai donc calculé la volatilité associé au chiffre d'affaires pour les entreprises dans les données INSEE et CRSP/Compustat, et représenté leur densité dans la Fig. d.2. On remarque alors qu'on peut calibrer raisonnablement bien ces données avec des lois inverses gamma, de densité

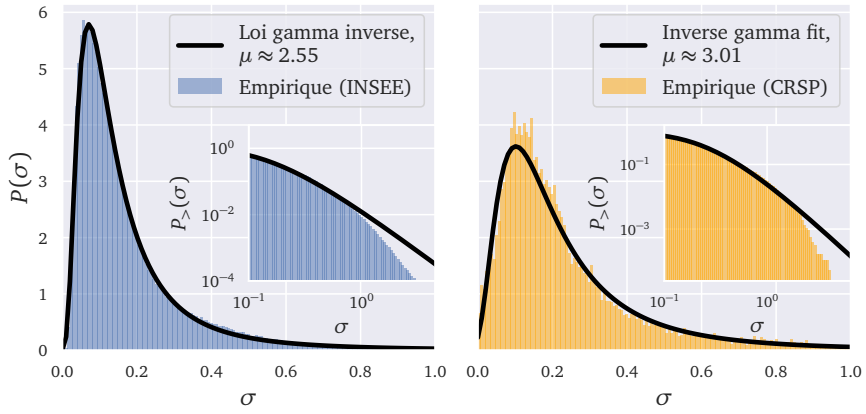


FIGURE d.2 : Distribution empirique des volatilités associés au chiffre d'affaires. Inséré : fonction de survie $P_>(\sigma) = \int_\sigma^\infty d\sigma' \, p(\sigma)$.

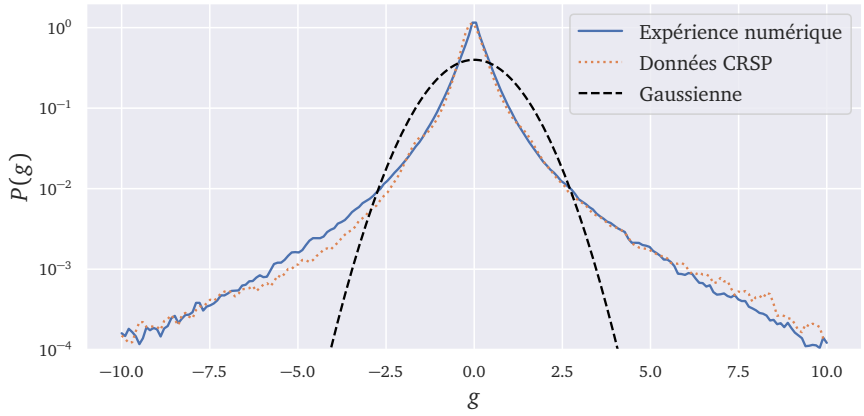


FIGURE d.3 : Expérience numérique : multiplication d'une variable gaussienne par la volatilité empirique de CRSP/Compustat. On voit alors qu'on reproduit bien la forme des taux de croissance observées.

$p_\sigma(\sigma) \propto \exp\left(-\frac{b}{(\sigma-a)}\right)\sigma^{-1-\mu}\mathbf{1}_{\sigma>0}$, où a et b contrôlent la forme de la distribution et μ la queue.

Il est alors évident que les taux de croissance sont trop hétérogènes pour être agrégés. Le lecteur trouvera dans la Sec. 6.3 une discussion détaillée de ce qui pourrait déterminer leur volatilité, mais il en ressort par exemple que la taille de l'entreprise ou son secteur d'activité sont des variables qui semblent influencer cela. On peut maintenant comprendre la raison du pic central et, peut-être, des queues.

Je propose l'expérience numérique suivante, qui consiste à définir $g = \sigma\tilde{g}$ où σ est tirée uniformément de l'ensemble de volatilités calculées empiriquement dans CRSP et où $\tilde{g}$ est pris comme une variable gaussienne. On remarque alors dans la Figure d.3 que cette expérience reproduit bien ce qu'on observe en réalité.

Je m'attarderai seulement sur la dépendance de la volatilité en la taille, puisque j'ai tenté de fournir une vérification empirique au mécanisme pro-

posé par Wyart et Bouchaud dans [279]. Leur argument consiste en effet à dire qu'une entreprise i est constituée d'un nombre K_i de sous-unités, chacune de taille s_α, $\alpha = 1, \ldots, K_i$. Les distributions des K_i et des s_α sont prises comme des lois de puissance, et donc une entreprise est l'agrégation d'entités invariantes d'échelle, et il en résulte alors notamment que cette structure est préservée si on "fusionne" deux entreprises. Ils arrivent ensuite à montrer que cette structure implique une relation volatilité/taille comme celle donnée en équation (d.3). J'ai voulu tester cela en étudiant comment cette statistique changeait en "fusionnant" des entreprises de façon aléatoire, car le modèle prédit que cette dépendance doit rester inchangée après cet exercice. Comme je le montre dans la Sec. 6.3, cela semble être le cas, ce qui donne du poids à cette explication.

d.2.1 *Caractère gaussien des taux de croissance*

Après cette étude de la volatilité, et après avoir montré que la raison qui explique la distribution des taux de croissance en forme de tente et avec des queues lourdes est probablement l'hétérogénéité de leurs volatilités, je propose ici une autre définition qui permet de comparer des taux de croissance d'entreprises différentes. En utilisant la technique dite de "leave one out" pour étudier des séries temporelles à volatilité hétérogène, donnée en [66], j'introduis alors le taux de croissance à échelle

$$\tilde{g}^i_{x,\Delta t}(t) = \frac{g^i_{x,\Delta t}(t) - \mathbb{E}_{t'}\left[g^i_{x,\Delta t}(t')\right]}{\sigma_{t'\neq t}\left[g^i_{x,\Delta t}(t')\right]}, \tag{d.7}$$

où $\sigma_{t'\neq t}$ correspond à la définition de l'équation (d.6) calculée en l'absence de l'observation en $t' = t$.

Ce point de vue est similaire à celui adopté pour étudier la statistique des variations de prix d'actions dans [134], et peut se comprendre en écrivant

$$g^i_{x,\Delta t}(t) = \mathbb{E}_{t'}\left[g^i_{x,\Delta t}(t')\right] + \sigma^i_{x,\Delta t}\tilde{g}^i_{x,\Delta t}(t), \tag{d.8}$$

c'est à dire en disant que le taux de croissance réalisé vient de fluctuations aléatoires modélisées par $\tilde{g}$ et amplifiées à une échelle fixée par σ.

En faisant cette mise à échelle on peut en effet se débarrasser de cette "amplification" due l'hétérogénéité des volatilités, aussi appelée hétéroscédasticité, et ne tenir compte que des fluctuations. Je trouve alors que les taux de croissance présentent une région centrale gaussienne qui s'agrandit au fur et à mesure que la période Δt s'élargit. Il subsiste des queues exponentielles étirées, mais il est difficile de savoir si leur présence est due à une estimation bruitée de la volatilité utilisée dans le dénominateur de l'équation (d.7) ou à une raison plus profonde. En m'inspirant de [55], j'introduis phénoménologiquement une densité de probabilité qui s'appa-

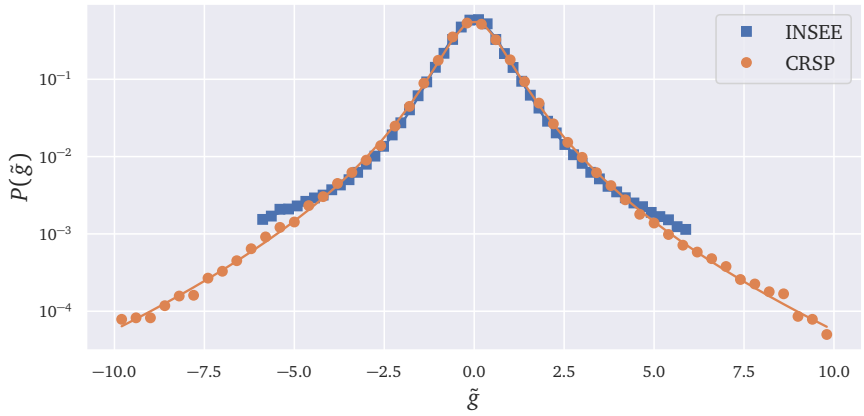


FIGURE d.4 : Taux de croissance à échelle $\tilde{g}$ du chiffre d'affaires dans les données CRSP/Compustat et INSEE. Les cercles et les carrés représentent les valeurs empiriques issues des données, alors que les lignes sont des calibrages non-paramétriques avec la densité définie en (d.9). Le calibrage fonctionne très bien, et les deux densités, bien qu'issues de données très différentes, semblent se superposer.

rente à la famille Subbotin, bien qu'avec une région centrale gaussienne au lieu de Laplace, donnée par

$$p_{\tilde{g}}(\tilde{g}) = C \exp\left(- \frac{(\tilde{g} - r_0)^2 / 2\sigma_0^2}{1 + \left(\tilde{g}/b_p\right)_+^{2-\mu_p} + \left(\tilde{g}/b_n\right)_-^{2-\mu_n}} \right), \tag{d.9}$$

avec C une constante de normalisation et $(x)_\pm = |x|\mathbf{1}(\pm x \geq 0)$ la partie positive (resp. négative).

Il est facile de voir que cette densité ressemble à une gaussienne à petit $\tilde{g}$, et que b_p et b_n contrôlent l'étendue de la région où cette approximation est valable, car en effet on peut voir que

$$p_{\tilde{g}}(\tilde{g}) \approx \begin{cases} C_0 \exp\left(-\frac{(\tilde{g}-r_0)^2}{2\sigma_0^2}\right) & \text{pour} \quad |\tilde{g}| \ll b_p, b_n \\ C_p \exp\left(-(b_p\tilde{g})^{-\mu_p}/(2\sigma_0^2)\right) & \text{pour} \quad \tilde{g} \gg b_p \\ C_n \exp\left(-(b_n\tilde{g})^{-\mu_n}/(2\sigma_0^2)\right) & \text{pour} \quad \tilde{g} \ll -b_n \end{cases}. \tag{d.10}$$

Le lecteur pourra constater en Fig. d.4 que les deux jeux de données ont des distributions très similaires et que la fonction proposée en équation (d.9) reproduit bien la forme observée.

On peut également s'intéresser à la convergence de $\tilde{g}$ vers une gaussienne quand Δt augmente, et ce pour différentes variables subjacentes x. Une façon très simple de faire cette étude est de calculer le troisième moment de la distribution conditionnellement au signe, c'est-à-dire d'étudier la quantité :

$$m_{3,\pm}(x, \Delta t) = \mathbb{E}\left(\tilde{g}^3_{x,\Delta t} \,\middle|\, \pm\tilde{g}_{x,\Delta t} > 0\right), \tag{d.11}$$

pour des différentes variables x et les retards Δt, et de comparer avec le résultat gaussien $m_{3,\pm} = \sqrt{\frac{2}{\pi}} \int_0^\infty \mathrm{d}x\, x^3 e^{-x^2/2} = \sqrt{\frac{8}{\pi}}$. Les résultats de cette analyse sont visibles en Fig. d.5, et la convergence est claire.

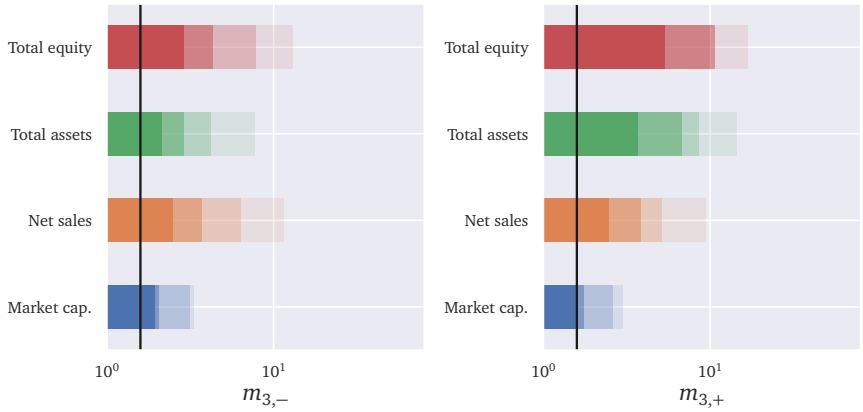


FIGURE d.5 : Moments $m_{3,\pm}$ pour des variables x différentes. L'intensité de couleur croissante correspond à $\Delta t = 1, 2, 4$ et 8 trimestres. La ligne noire continue correspond à la valeur gaussienne. On voit que la queue droite de la capitalisation boursière ("market cap.") converge très rapidement vers une gaussienne, alors que la queue gauche le fait plus lentement. Au total, toutes les variables montrent une convergence vers une distribution gaussienne, même si cela se produit de manière asymétrique et à des vitesses variables pour chaque variable. Les noms des variables sont en anglais, mais indiquent, du haut vers le bas, le total des fonds propres, le total des actifs, le chiffre d'affaires et la capitalisation boursière.

Une analyse similaire, qui montre que les paramètres b_p et b_n croissent avec Δt, est également faite en Sec. 6.4.

d.3 DYNAMIQUE DES TAUX DE CROISSANCE

L'idée d'introduire les variables $\tilde{g}$ était de se débarrasser de l'amplitude idiosyncratique σ de chaque firme et d'obtenir des fluctuations de taille comparable, de sorte qu'on peut les comparer les unes avec les autres. Ainsi, $\tilde{g}$ contient toutes les informations des fluctuations qu'une entreprise a subies.

On peut aller étudier aisément le lien entre des fluctuations à t et à $t + \tau$ en calculant des corrélations. Puisqu'on a des séries très courtes ceci devrait être impossible, mais la stratégie que j'adopte consiste en fait à calculer des estimations bruitées de chaque corrélation, entreprise par entreprise, puis de moyenner ces corrélations sur l'ensemble des firmes de mon échantillon. Ceci est différent de ce qu'on peut faire en finance, où les séries sont beaucoup plus longues. Pour cette partie, l'unité de temps est fixée au trimestre.

Pour une entreprise i, je définis alors la corrélation décalée entre deux quantités x et y comme

$$C^i_{x,y}(\tau) = \mathbb{E}_t\left[\tilde{g}^i_x(t+\tau) g_y(t)\right], \tag{d.12}$$

où la définition est transparente, et la moyenne n'est prise que sur le temps. Je définis ensuite la corrélation globale entre les deux quantités comme

$$C_{x,y}(\tau) = \mathbb{E}_i\left[C^i_{x,y}(\tau)\right]. \tag{d.13}$$

Une autre quantité utile dans cette analyse est la *signature de volatilité*. Pour la comprendre, il faut d'abord définir le variogramme : j'invite le lecteur à imaginer une série temporelle $x^i(t)$ décrivant l'évolution d'une quantité x au cours du temps pour une entreprise i. Si on réutilise la définition du taux de croissance, alors on peut écrire

$$\log\big(x^i(t)\big) = \log\big(x^i(0)\big) + \sum_{k=0}^{t-1} g^i_x(k), \tag{d.14}$$

et on définit ensuite le variogramme comme

$$\mathscr{V}^i(\tau) = \mathbb{E}_t\left[\log\big(x^i(t+\tau)/x(t)\big)^2\right] - \left(\mathbb{E}_t\left[\log\big(x^i(t+\tau)/x(t)\big)\right]\right)^2, \tag{d.15}$$

qu'on peut également écrire comme

$$\mathscr{V}^i_x(\tau) = \left(\sigma^i_x\right)^2\left(\tau + \sum_{k=1}^{\tau}(\tau-k)\,C_{x,x}(k)\right). \tag{d.16}$$

Il est aisé de remarquer que si $\log\big(x^i\big)$ suit une marche purement aléatoire alors le variogramme se comporte comme $\mathscr{V}(\tau) \propto \tau$, alors que pour une quantité déterministe $x(t) \propto t$ le variogramme est plutôt $\propto \tau^2$. De même, une série d'incréments anti-corrélés présente un variogramme sous-linéaire. Ceci motive l'introduction de la signature de volatilité,

$$v^i_x(\tau) = \sqrt{\frac{\mathscr{V}^i_x(\tau)}{\left(\sigma^i_x\right)^2\tau}} = \sqrt{1 + \sum_{k=1}^{\tau}\left(1-\frac{k}{\tau}\right)C_{x,x}(k)}, \tag{d.17}$$

qui est croissante quand les incréments sont corrélés positivement, plate quand ils sont décorrélés, et décroissante quand ils sont anti-corrélés. Les résultats de cette analyse sont visibles en Fig. d.6, et montrent notamment que la dynamique suivie par le prix de marché d'une entreprise et par ses données fondamentales sont différents. Pour le premier on retrouve un phénomène bien connu (voir par exemple [60] et les explications proposées dans [86, 174]), où il y a essentiellement des phénomènes de suivi de tendance (c'est à dire qu'une augmentation est probablement suivie d'une autre augmentation) sur des échelles de l'ordre de l'année, suivies de phénomènes de retour à la moyenne (le cas inverse au précédent, une augmentation est suivie d'une diminution modératrice). On retrouve globalement des mécanismes similaires pour les autres quantités, mais moins marqués et sur des échelles de temps différentes.

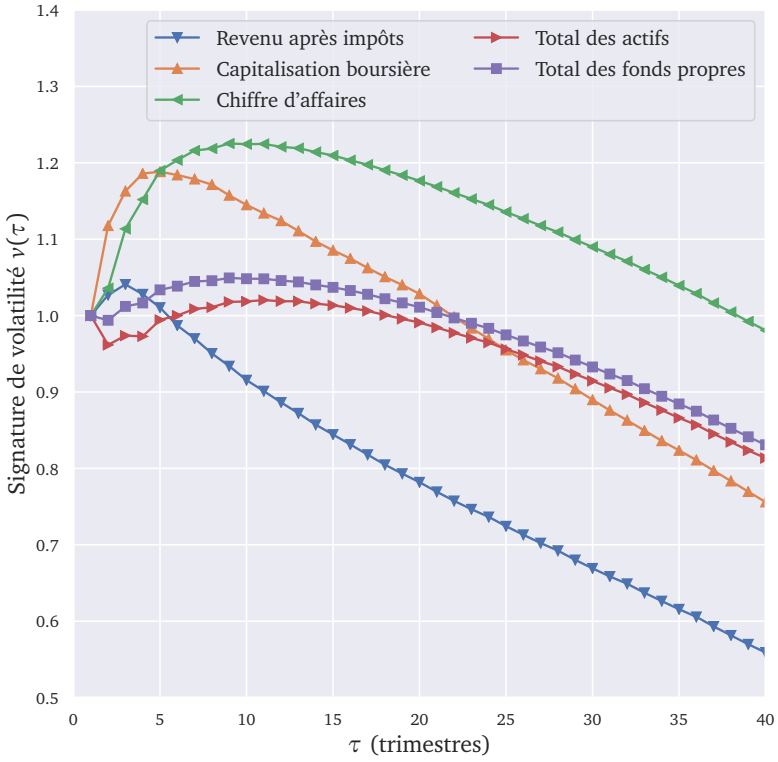


FIGURE d.6 : Signature de volatilité calculée pour différentes grandeurs x dans les données CRSP/Compustat. Les courbes sont croissantes quand il y a des mécanismes de renforcement de la tendance et sont décroissantes quand il y a des mécanismes de retour à la moyenne.

d.4 CONCLUSION

Le travail présenté dans cette section est toujours en cours et constitue le cœur du travail empirique que j'ai effectué au cours de ma thèse. La contribution la plus importante est, je crois, la prise en compte de l'hétéroscédasticité de la croissance des entreprises. Cela peut être interprété comme une nouvelle version de la loi de Gibrat, dans un sens plus faible que celui où l'on suppose la statistique indépendante de l'entreprise, en affirmant que les taux de croissance peuvent être écrits comme

$$g^i_{x,\Delta t} = m_i + \sigma^i_{x,\Delta t}\tilde{g}, \tag{d.18}$$

où le taux de croissance moyen m_i et sa volatilité σ dépendent entièrement de l'entreprise et de la quantité x considérée, mais où la variable $\tilde{g}$ est considérée ayant une statistique indépendante de la firme.

Il est donc possible de poursuivre dans ce sens. Il serait ainsi intéressant de se demander, par exemple, précisément comment les paramètres m et σ dépendent des caractéristiques de l'entreprise. Un autre indice intéressant

consiste à étendre l'analyse des corrélations intra-entreprise effectuée en Sec. 6.5 aux corrélations interentreprises, et à essayer de montrer comment les dépendances de la *supply-chain* peuvent devenir des fluctuations des ventes et de la production, sur le modèle du projet développé en détail dans le Chapitre 5. Les bases de données avec lesquelles j'ai travaillé permettent d'explorer ces questions, et je suis persuadé que l'analyse développée ici portera davantage de fruits.

Ceci conclut le travail que j'ai effectué pour modéliser et étudier les entreprises et leur croissance, même si cela m'a amené à poser d'autres questions qui restent ouvertes. Il s'agit, pour n'en citer que quelques-unes, de la question de l'existence d'un équilibre économique, qui est soulevée dans le Chapitre 5, et de la manière dont on peut y converger ; de la question de savoir comment des interactions microscopiques au niveau des agents peuvent conduire à des phénomènes collectifs totalement imprévisibles, comme l'a développé l'étude des avalanches dans le Sec. 3.4.2, et enfin l'impact de l'hétérogénéité et la possibilité d'avoir des fluctuations endogènes, tels que soulevés plus largement dans cette partie de ma thèse.

La partie suivante est une tentative d'étudier certaines de ces questions. Après ce que l'on pourrait qualifier d'étude macro-économique sur les entreprises, il peut sembler un peu trop ambitieux d'envisager des choses telles que le comportement de foules, l'exploitation de zones de pêche et les effets de mémoire dans la théorie de la décision, mais je crois fermement qu'il est nécessaire d'aborder ces problèmes pour avoir une perspective meilleure et plus pertinente de la modélisation des phénomènes économiques et sociaux.

Je cite, au début du Chapitre 4, un certain nombre d'économistes qui ont estimé que l'arsenal théorique dont ils disposaient au moment de la crise des années 2007—2008 n'était pas adapté pour faire face à une situation aussi difficile. Cela a conduit à une remise en question agressive d'un ensemble d'axiomes très présents dans la théorie économique dominante [62], tels que le paradigme de l'agent représentatif rationnel, l'ergodicité ou la notion même d'équilibre. Cela suggère de regarder au-delà de l'infiniment sage et rationnel *homo economicus* pour développer des modèles plus réalistes pour l'homo sapiens. C'est, en résumé, le programme défini par l'économie de la complexité [40] qui a motivé la partie qui suit, la dernière de ma thèse.

e

ERGODICITÉ, MÉMOIRE, IMITATION ET SYSTÈMES COMPLEXES

Pour modéliser le comportement global des individus, l'économie classique a souvent recours à la vision de l'*homo œconomicus*, un agent parfaitement rationnel dont le comportement vise à maximiser une fonction d'utilité. En effet, cette vision permet souvent d'avoir des modèles qui sont solubles analytiquement, et de fournir ce que les économistes appellent des "micro-fondations", c'est-à-dire des justifications du comportement à l'échelle de l'individu, bien que le paradigme de la rationalité ait fait l'objet de vives critiques ces dernières années. Cette vision s'identifie aussi souvent à une position dite libérale, où on estime que l'économie peut s'auto-organiser dans un état qui est qualifié d'optimal. Cependant, comme le dit bien Kirman, cette vision ne tient pas face à un examen empirique rigoureux, ou même lorsque confrontée à des modèles analytiques plus sophistiqués [158].

En effet, les économies modernes sont constituées de millions d'agents, souvent organisés en groupes ou entreprises, qui interagissent ensemble à travers leurs relations sociales et aussi par un marché. Les agréger en une seule entité revient à jeter le bébé avec l'eau du bain, car on néglige alors la façon dont ils interagissent, et donc comment ils peuvent se coordonner ou s'entraver. La tentation de déclarer la mort de l'agent représentatif est donc assez forte [154].

Dans les parties précédentes, le lecteur s'est déjà familiarisé avec le modèle RFIM (Random Field Ising Model) : dans ce modèle, les agents peuvent faire un choix binaire ± 1, ou A et B, mais ils subissent l'influence de leurs pairs et d'un "champ externe" qui capture toute l'information publique. Les variations de ce champ externe peuvent alors provoquer des "avalanches d'opinions" à grande échelle, où un agent change d'avis et influence les autres pour qu'ils fassent de même. En particulier, ces phénomènes collectifs spectaculaires, similaires aux transitions de phase en physique, comme la transition liquide/solide pour l'eau, ne se produisent qu'en raison des interactions. C'est déjà un grand saut conceptuel, qui est bien résumé dans le manifeste anti-réductionniste d'Anderson, *More is different* [17].

Une autre idée profondément ancrée dans cette discipline, étroitement liée à celle de l'agent représentatif, est celle de l'équilibre économique. Cette idée a, dans les termes de Farmer et Geanakoplos [109], "un impact énorme sur la façon dont les économistes pensent, et définit en fait à bien des égards la façon de penser des économistes". Dans ce cadre réduit, et comme le souligne le chapitre 4, les fluctuations des valeurs d'équilibre ne peuvent être que de nature exogène, c'est-à-dire qu'elles proviennent d'une "main de Jupiter" providentielle, pour citer Adam Smith. Cependant,

un nombre croissant de voix demande à prendre en compte les origines endogènes des fluctuations [63, 245].

Ainsi, je présenterai dans cette partie le modèle de recrutement de Kirman et Föllmer [1]. Il s'agit d'un modèle stochastique avec de l'auto-renforcement, assez similaire aux urnes de Pólya, qui permet d'obtenir des avalanches d'opinion d'origine purement endogène.

Ma contribution à ce modèle et à cette littérature consiste à fournir, avec Fosset, Benzaquen et Bouchaud, une solution analytique complète, donnant ainsi un aperçu de la manière dont les avalanches se produisent. Cette solution a fait l'objet d'une publication, disponible dans la référence [189]. D'autres travaux réalisés avec Fosset, Benzaquen et Kirman s'inspirent largement de ce modèle pour expliquer des faits stylisés empiriques décrivant la dynamique de bateaux de pêche en mer Adriatique. Cette dernière étude n'a pas encore été publiée, mais les principaux résultats sont également résumés ci-dessous.

Enfin, je présenterai également une autre contribution, dont l'idée à d'abord été esquissée par Bouchaud dans [63] et qui vise à aller au delà des modèles classiques de théorie de la décision par un biais autre que celui des effets d'imitation. Il s'agit en effet de tenir compte de possibles effets de mémoire dans les choix des individus. La résolution de ce modèle, encore très simple et épuré, a mené à une publication [192], où nous avons montré qu'il mène par ailleurs à une phénoménologie intéressante de "super-vieillissement", avec des applications possibles dans la théorie des gels et des mousses.

e.1 DESCRIPTION DU MODÈLE

Le *modèle des fourmis de Kirman* [157] figure sans aucun doute parmi les modèles-jouets les plus connus de la littérature en économie comportementale. Bien qu'inspiré initialement par l'expérience décrite ci-dessous, ses conclusions ont des implications bien au-delà du comportement collectif des insectes, puisqu'il a été utilisé pour modéliser les changements de sentiment des agents économiques, le renversement de tendance sur les marchés financiers, l'élevage et l'influence sociale, etc. Il s'apparente également à un autre modèle célèbre de la dynamique de populations avec des espèces concurrentes : le modèle de Moran [193].

Des entomologistes ont d'abord été surpris en voyant des fourmis qui, pouvant choisir entre deux sources de nourriture identiques et inépuisables A et B, avaient tendance à se concentrer sur l'une d'entre elles pendant un certain temps, mais passaient parfois collectivement à l'autre sans raison apparente[98, 39]. Ce comportement intermittent est très similaire à celui observé chez les humains qui choisissent entre des restaurants équiva-

1 Bien que Föllmer ne figure pas dans la liste des auteurs, c'est lui qui a fourni la solution mathématique au modèle proposé par Kirman. D'après le récit de Kirman, Föllmer n'a pas jugé sa contribution au modèle, c'est à dire sa résolution mathématique à l'état stationnaire, pertinente au départ et n'a donc pas signé l'article. À la demande de Kirman suite à la publication de Ref. [189], j'attribue le modèle à ces deux personnes.

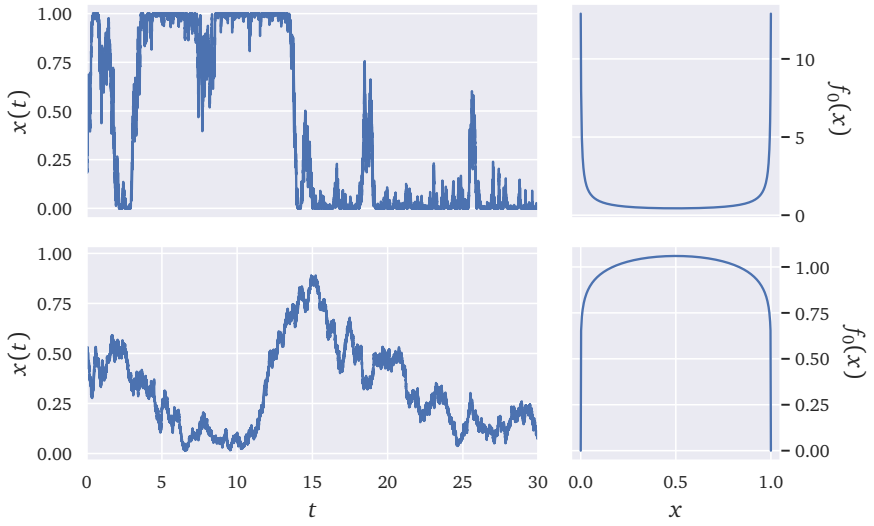


FIGURE e.1 : Simulations du modèle dans la limite continue. Les deux graphes du haut correspondent à $\alpha = 0.1 < 1$, alors que ceux du bas correspondent à $\alpha = 2 > 1$. Les graphes à gauche représentent l'évolution de $x(t)$, donnée par l'équation (e.3), et ceux de droite montrent la distribution stationnaire $f_0(x)$.

lents [38], ou sur les marchés financiers [230, 239, 171], et est conforme à l'idée que la plupart des grandes fluctuations sont de nature purement endogène. Pour rendre compte de ce comportement complexe, Kirman a proposé un modèle très simple [157] basé sur le recrutement en tandem.

Considérons N fourmis et désignons par $k(t) \in [0, N]$ le nombre de fourmis se nourrissant de la source A à l'instant t. Le modèle prescrit que lorsque deux fourmis se rencontrent, l'une d'elles convertit l'autre avec une probabilité de μ/N, mais chaque fourmi peut aussi changer d'avis spontanément avec une probabilité ε. Sans rentrer dans les détails des calculs, décrits en détail dans le Chapitre 7, il est possible de montrer que la densité $f(x,t)$ de la variable $x(t) = k(t)/N$ suit l'évolution déterminée par l'équation de Fokker-Planck suivante :

$$\partial_t f = \partial_x J^f, \quad \text{avec} \quad J^f(x,t) = -\varepsilon(1-2x)f(x,t) + \mu\partial_x\left[x(1-x)f(x,t)\right], \tag{e.1}$$

où cette dernière quantité J^f représente le flux de probabilité de x vers $x + \mathrm{d}x$, qui doit s'annuler en $x = 0$ et $x = 1$.

On peut alors trouver facilement l'état stationnaire en imposant la nullité du flux, $J^f = 0$, ce qui donne un état stationnaire

$$f_0(x) = \frac{\Gamma(2\alpha)}{\Gamma^2(\alpha)}\left[x(1-x)\right]^{\alpha-1}, \quad \text{avec} \quad \alpha := \frac{\varepsilon}{\mu}, \tag{e.2}$$

en accord avec le résultat de Föllmer et Kirman.

Il est donc aussi possible de trouver l'équation différentielle stochastique suivie par x, c'est-à-dire

$$\dot{x} = \varepsilon(1-2x) + \sqrt{2\mu x(1-x)}\eta(t)\,, \tag{e.3}$$

avec η un bruit blanc gaussien de variance unitaire. On peut alors constater que le terme déterministe $\varepsilon(1-2x)$ atteint son maximum sur les bords 0 et 1 et tend à ramener x vers $1/2$, alors que le terme stochastique a l'effet opposé. Le terme diffusif est en effet proportionnel à $\sqrt{2\mu x(1-x)}$, et est donc maximal en $x = 1/2$ et tend donc à écarter le système de cette valeur.

On est alors tenté de se débarrasser de la dépendance en x devant le terme diffusif. Cela justifie le changement de variable

$$\sin\varphi = 2x-1. \tag{e.4}$$

On peut alors montrer que la densité $g(\varphi,t)$ de cette nouvelle variable suit l'équation de Fokker-Planck

$$\begin{aligned} \partial_t g = \mu\partial_\varphi J^g, \quad \text{avec} \quad J^g(\varphi,t) = 2\beta\tan\varphi\, g(\varphi,t) + \partial_\varphi g(\varphi,t)\,, \\ \text{et} \quad \beta := \alpha - \frac{1}{2}\,, \end{aligned} \tag{e.5}$$

dont le flux doit vérifier $J^g(\pm\pi/2,t) = 0$ à tout instant t. On peut maintenant poser un changement de variable classique, utilisant les méthodes dites à énergie en résolution d'équations aux dérivées partielles, et poser

$$g(\varphi,t) := \sqrt{g_0(\varphi)}\Psi(\varphi,t)\,, \tag{e.6}$$

où $\Psi(\varphi,t) \to \sqrt{g_0(\varphi)}$ quand $t \to \infty$.

On obtient alors une équation de Schrödinger pour Ψ,

$$-\frac{1}{\mu}\partial_t\Psi = \mathbf{H}\Psi, \tag{e.7}$$

où l'hamiltonien $\mathbf{H}$ est défini comme :

$$\mathbf{H} := -\partial_{\varphi\varphi} + V(\varphi), \quad V(\varphi) := -\beta + \beta(\beta-1)\tan^2\varphi, \tag{e.8}$$

et avec des conditions aux limites

$$\left[\cos^\beta\varphi\big(\beta\tan\varphi\Psi(\varphi,t) + \partial_\varphi\Psi(\varphi,t)\big)\right]_{\varphi=\pm\pi/2} = 0\,. \tag{e.9}$$

Il s'agit ici d'un potentiel connu sous le nom de potentiel de Pöschl-Teller, qui a été résolu déjà dans [201, 261], mais avec des conditions aux limites $\Psi(\pm\pi/2,t) = 0$. On peut vérifier en fait que la solution de [261] vérifie les conditions aux limites que j'ai imposées ici, et on obtient alors une solution complète :

$$\Psi(\varphi,t) = \lambda_0\sqrt{g_0(\varphi)} + \sum_{n>1}\lambda_n\Psi_n(\varphi)e^{-\mu\mathscr{E}_n t}, \tag{e.10}$$

avec les états propres

$$\begin{aligned}\mathscr{E}_n =& n(2\alpha+n-1)\,,\\ \Psi_{2k}(\varphi) =& A_{2k}(\beta)\ {}_2F_1\left(-k,\beta+k;\beta+\tfrac{1}{2},\cos^2\varphi\right)\cos^\beta\varphi\,,\\ \Psi_{2k+1}(\varphi) =& A_{2k+1}(\beta)\ {}_2F_1\left(-k,\beta+k+1;\beta+\tfrac{1}{2},\cos^2\varphi\right)\sin\varphi\ \cos^\beta\varphi\,,\end{aligned} \tag{e.11}$$

où ${}_2F_1$ est la fonction hypergéométrique.

On résout alors aisément le modèle en revenant à $g(\varphi,t)$ puis à $f(x,t)$, et on obtient

$$f(x,t) = f_0(x) + \sum_{n>1} e^{-\mu\mathscr{E}_n t}\Psi_n(\varphi_0)f_n(x)\,, \tag{e.12}$$

où

$$f_n(x) = \frac{\sqrt{g_0(\varphi(x))}\,\Psi_n(\varphi(x))}{2\sqrt{x(1-x)}}\,. \tag{e.13}$$

En particulier, on voit que le temps de convergence à l'état stationnaire, donné par la plus petite énergie non nulle $\mathscr{E}_1$, est proportionnel à $1/2\varepsilon$, ce qui implique que les avalanches sont provoquées par une seule fourmi qui décide spontanément de changer d'avis et finit par entrainer toutes les autres. C'est le résultat principal de la solution du modèle, et clôt donc ce résumé.

e.1.1 *Prolongation au modèle de poissons*

La partie 7.2 présente une extension du modèle des fourmis. Pour résumer, nous avons, avec Antoine Fosset, Michael Benzaquen et Alan Kirman, étudié la dynamique de chalutiers en mer Adriatique, avec des données issues du Global Fishing Watch [133]. On retrouve dans ces données des faits stylisés qui rappellent le modèle des fourmis : des distributions en loi bêta et des corrélations exponentielles.

Nous avons donc étendu le modèle des fourmis, en comparant la colonie de fourmis à un port de pêche, et les sources de nourriture à des zones de pêche qui deviennent maintenant épuisables. Comme modèle simple, nous avons prescrit une évolution du type Lotka-Volterra pour la population de poissons dans une zone de pêche, c'est à dire qu'en absence de pêcheurs il y a une croissance logistique jusqu'à une certaine valeur de saturation, définissant la capacité de la zone, alors que quand des bateaux sont présents il y a diminution de la population à cause de la pêche. La probabilité de changement spontané, ε dans le modèle précédent, dépend alors de la population de chaque zone de pêche, et augmente lorsque le nombre de poissons dans la zone diminue.

Le modèle ainsi décrit devient relativement compliqué, mais peut être résolu dans une approche de *champ moyen*, où on remplace de façon auto-cohérente la population de poissons par sa moyenne temporelle. On peut

alors tout calculer, comme dans le cas du modèle de fourmis, et reproduire les faits stylisés décrits plus haut.

e.2 EFFETS DE MÉMOIRE

Comme mentionné précédemment, dans la théorie du choix rationnel, les individus fixent leurs choix en maximisant une fonction d'utilité. Chaque choix qu'un individu peut faire se voit ainsi attribuer une certaine "utilité", c'est-à-dire une quantité mesurant la satisfaction qu'il procure à l'agent. Ce cadre est souvent accompagné de l'hypothèse que l'agent considère *tous* les choix disponibles qui lui sont présentés, pèse leurs utilités les unes par rapport aux autres, puis fait son choix en tenant compte des contraintes éventuelles, telles qu'un budget limité.

Cette vision du comportement humain a fait l'objet d'un certain nombre de critiques, avec par exemple Simon [241] soulignant que les individus peuvent chercher à être satisfaits plutôt que de faire une optimisation complète, puisqu'il existe à la fois un coût informationnel et un biais cognitif lié à la prise en compte de l'univers des choix disponibles. Trouver l'optimum d'une fonction d'utilité peut être parfois un problème de calcul si difficile que même les ordinateurs les plus puissants ne seraient pas capables de le résoudre en un temps raisonnable. C'est ainsi qu'est née l'idée de la rationalité limitée comme moyen de modéliser les agents réels [243, 233, 22, 129].

Une idée intéressante développée dans [147] est de dire que l'utilité associée à un certain choix peut également dépendre de notre mémoire si ce choix a déjà été fait dans le passé. Le modèle proposé ci-dessous résume cette idée, et je montrerai que cela aussi peut conduire à des choix "sous-optimaux", qui ne sont faits que par désir d'imitation du passé. Il s'agit de la "formation d'habitudes" en économie [78, 1, 92, 79, 116, 219]. Ainsi, les effets de mémoire modifient le paysage d'utilité d'une manière qui peut rendre des choix objectivement sous-optimaux subjectivement optimaux. Dans le cas d'une mémoire à durée suffisamment longue, les agents peuvent, de manière auto-réalisatrice, se "piéger" dans un certain choix et cesser d'explorer des choix alternatifs. Par ailleurs, ce mécanisme d'auto-réalisation brise complètement l'hypothèse d'ergodicité.

e.2.1 *Description du modèle*

Prenons N choix $(x_i)_{1\leq i\leq N}$, auxquels on attribue une utilité

$$U(x_i,t) = U_0(x_i)\left(1+\sum_{t'=0}^{t}\phi(t-t')\mathbf{1}_{x(t')=x_i}\right), \tag{e.14}$$

où le premier terme à droite est l'utilité intrinsèque ou objective associée au choix, tandis que le second tient compte des effets de mémoire, qui modifient l'utilité de ce choix pour la seule raison que l'individu l'a choisi

dans le passé.[2] Le noyau de mémoire ϕ est décroissant, et encode le fait que les choix plus récents ont un effet plus fort sur l'utilité, et $x(t)$ indique le choix de l'individu au moment t. Ainsi, l'histoire passée "cisèle" le paysage d'utilité, d'une manière similaire aux fourmis qui laissent une trace de phéromones qui guide les autres fourmis sur le même chemin, ou les rivières qui créent leur propre lit par érosion. Dans ce modèle, je ne considère que le cas $\phi > 0$ où le noyau *renforce* les choix antérieurs.

Ce modèle est ensuite muni de la dynamique suivante. Un individu, ayant fait le choix x_i à l'instant t, tire un choix alternatif x_j d'un certain ensemble de "choix proches" ∂_i, par exemple l'ensemble des voisins de i dans un graphe $\mathcal{G}$, avec probabilité :

$$\mathcal{T}_{x_i \to x_j} = \frac{\mathbf{1}_{x_j \in \partial_i}}{\mathcal{N}_i}, \quad \mathcal{N}_i := \sum_j \mathbf{1}_{x_j \in \partial_i}, \tag{e.15}$$

où $\mathcal{N}_i$ est le nombre de voisins de i. Le cadre que je présente est assez polyvalent puisque la topologie du graphique $\mathcal{G}$ est arbitraire, mais dans ce résumé je me limite au cas d'un graphe complet, où tous les choix sont accessibles.

Le choix cible x_j est ensuite adopté avec la règle logit, très courant en théorie du choix [18] :

$$p(x_i \to x_j) = \frac{1}{1 + e^{\beta[U(x_i,t) - U(x_j,t)]}}, \tag{e.16}$$

où β est l'"intensité" et explique le degré de rationalité (analogue de la température inverse en mécanique statistique). En effet, tant que $0 < \beta < \infty$, l'agent est susceptible lorsque $U(x_j, t) > U(x_i, t)$ (comportement d'optimisation), mais la probabilité de choisir un choix ayant une utilité moindre est non nulle, ce qui modélise la rationalité limitée. Dans la limite $\beta \to 0$ (équivalent à la limite de température infinie en physique), l'agent explore tout l'espace des choix possibles sans tenir compte de leur utilité. Dans la limite opposée $\beta \to \infty$ (ou température zéro) l'agent ne passe qu'à des choix d'une utilité plus élevée, mais cela implique également qu'il peut rester dans un maximum local au lieu de prendre la chance d'explorer toutes les possibilités disponibles.

L'ajout du noyau introduit la possibilité que l'agent soit piégé dans un choix "sous-optimal" à cause des effets de mémoire : s'il reste un temps suffisamment long sur un choix, l'utilité associée à ce dernier croîtra et donc augmentera la probabilité que l'agent y reste. Pour étudier les effets de la mémoire seule, je considère seulement le cas homogène $U_0(x_i) = 1, \ \forall i$.

La clef pour attaquer ce modèle consiste à calculer la probabilité $P_>(\tau)$ qu'un agent soit resté sur un même site i en $t = 0, \ldots, \tau$, c'est-à-dire qu'il soit piégé au moins pendant un temps τ. Cette probabilité se calcule

2 On peut aussi penser, dans le langage du physicien, à un paysage énergétique (c'est-à-dire moins l'utilité) où l'énergie d'un site ou d'une configuration donnée augmente ou diminue si le système a déjà visité ce site.

aisément, comme le produit des probabilités de ne pas quitter ce choix à ces temps, dont l'expression est

$$p_{\text{rester}}(t) = 1 - \sum_{j \in \partial_i} \mathscr{T}_{x_i \to x_j} \frac{1}{1 + e^{\beta[U(x_i,t) - U(x_j,t)]}}. \tag{e.17}$$

Cette expression se simplifie pour un graphe complet, et en notant $\Phi(t) = U_0 \sum_0^t \phi(s)$ on peut alors écrire

$$P_>(\tau) = \prod_{t=0}^{\tau} \left[1 + e^{-\beta\Phi(t)}\right]^{-1} \approx e^{-I(\tau)}, \tag{e.18}$$

avec $I(\tau) := \int_0^\tau \mathrm{d}t \, \ln\left[1 + e^{-\beta\Phi(t)}\right]$. Cette équation décrit alors la statistique du temps de piégeage. Il est ensuite possible de classifier les différents comportements du modèle selon le comportement de Φ.

e.2.2 *Mémoire courte*

En effet, si $\lim_{t\to\infty} \Phi(t) = \|\phi\| < +\infty$, alors $I(\tau) \approx \lambda\tau$ for $\tau \to \infty$, avec

$$\lambda := \ln\left[1 + e^{-\beta\|\phi\|}\right]. \tag{e.19}$$

Cela signifie que la distribution des temps de piégeage a une décroissance exponentielle, avec un temps moyen de piégeage de l'ordre de $1/\lambda$. On peut donc avoir l'impression d'être piégé en observant le système seulement pendant une durée T plus petite que $1/\lambda$, le temps ergodique, mais on explore tous les choix dans la limite $T \to \infty$.

e.2.3 *Mémoire longue*

On suppose alors que $\Phi(t)$ diverge, ce qui est possible par exemple avec des noyaux en loi de puissance,

$$\phi(t) = \frac{C}{(1+t)^\gamma}. \tag{e.20}$$

Le cas $\gamma > 1$ nous ramène au cas précédent, et il faut donc étudier le cas $\gamma \leq 1$. Quand $\gamma < 1$, on trouve que $\Phi(t) \propto t^{1-\gamma}$ pour t grand. Par conséquent, $I(\tau)$ converge vers une limite finie I_∞ à grand τ. Cela signifie qu'il y a une probabilité finie $P_\infty = e^{-I_\infty}$ que le choix soit fait pour toujours.

Quand $\gamma = 1$, $\Phi(t) \approx CU_0 \ln t$ à grand t. Cela conduit à étudier trois autres sous-cas :

1. Quand $\beta C U_0 > 1$, $I(\tau)$ converge à nouveau vers une limite finie quand $\tau \to \infty$, et donc on est piégé pour toujours.

2. Au point de transition, défini par $\beta_c^\star = (CU_0)^{-1}$, on trouve que $P_>(\tau)$ décroît comme τ^{-1}, et donc la distribution des temps de piégeage

est une loi de Zipf, en τ^{-2}. Il s'agit du cas marginal de vieillissement qu'on peut trouver fréquemment dans la littérature [45, 14]. Pour un temps d'observation fini T, le temps moyen de piégeage croît comme $\ln T$.

3. Quand $\beta C U_0 < 1$, $I(\tau)$ se comporte à grand τ comme $\exp\big(-\tau^b/b\big)$, avec $b = 1-\beta C U_0 > 0$. Le temps moyen de piégeage $\langle\tau\rangle$ est donc fini. On peut montrer également que $\langle\tau\rangle$ diverge en b^{-1} lorsque $b \to 0$, mais des moments d'ordre plus élevé, $\langle\tau^k\rangle$ avec $k > 1$, divergent plus vite, comme $\exp\big((k-1)/b\big)$, selon la loi dite de Vogel-Tamman-Fulcher, voir [97].

e.2.4 *Vieillissement*

On peut en fait être un peu plus précis sur la signification de l'auto-piégeage pour un temps d'observation T fini lorsque $\beta > \beta_c^\star$. En fait, le système *vieillit*, dans le sens suivant [59, 188] : supposons que le choix de l'agent au moment T est un certain x_i et demandons-nous : Quelle est la probabilité $\mathscr{P}(t, T)$ que l'agent n'ait jamais changé d'avis entre T et un moment ultérieur $T + t$? Dans la phase d'exploration libre $\beta < \beta_c^\star$, cette probabilité est indépendante de T pour $T \gg 1$: le processus est invariant dans le temps. Dans la phase piégée $\beta > \beta_c^\star$, $\mathscr{P}(t, T)$ peut être estimé. Le résultat prend la forme classique de vieillissement,

$$\mathscr{P}(t, T) \approx \exp\left(\frac{1}{a(T+t)^a} - \frac{1}{aT^a}\right), \quad a = \frac{\beta}{\beta_c^\star} - 1 > 0. \tag{e.21}$$

Dans le régime $t \ll T$, $\mathscr{P}(t, T)$ est une fonction de t/T^{1+a} (en fait $1-\mathscr{P}(t, T) \approx t/T^{1+a}$), un régime appelé *super-aging/super vieillissant* [61], puisque le temps nécessaire à changer d'avis croît comme $T^{\beta/\beta_c^\star}$, et donc plus vite que l'age T du système. C'est un cadre très intéressant, car nous ne connaissons pas de modèles simples conduisant à un comportement de vieillissement super-linéaire. Par conséquent, les effets de mémoire du type dont il est question ici pourraient très bien être une piste intéressante pour interpréter des expériences comme celles rapportées dans [87].

FOREWORD

Statistical physics and *macroeconomic fluctuations* are two fields that don't seem to have anything in common. To the layman, at least in my experience, the word *physics* evokes the study of celestial bodies, subatomic particles, lasers or other objects that seem far removed from notions such as their country's gross domestic product or unemployment rate, and I have often been asked how these two subjects can be related. My answer usually begins with the clarification of what *statistical* physics is: a discipline that strives to understand the collective behaviour of a large number of entities, an enterprise whose modern origins can be found in Boltzmann and Maxwell's kinetic theory of gasses and that has since shed light on the mechanisms behind phenomena ranging from the paramagnetic/ferromagnetic phase transition to the surprising collective cohesion of a flock of starlings.

The etymological origin of the word *statistics*, showing the initial preoccupation of the discipline with the *State*'s affairs, also hints at another historical fact: Maxwell claimed that his kinetic theory was indebted to the work of Adolphe Quetelet [221], a man who was fascinated by the regularities he saw in aggregated data describing societies and who tried to create what he called "social physics" [54]. In that sense at least, it seems that statistical physics was originally influenced by the attempt to understand large-scale phenomena in society rather than the other way around. The emergence of the field of *Complex Systems*, and also of what has been now dubbed *"Econophysics"*, is therefore rooted in the history of statistical physics, and it seems only natural that more recent findings within that discipline should also spill over into the study of socio-economic phenomena.

It is with this in mind that I have decided to study *anomalous macroeconomic fluctuations* as a physicist. The central question that has interested me, and that shall be repeated a number of times in this document, can be summarized as the attempt to understand how the different parts that make up the economy, from firms down to individual agents, interact and become an aggregate whose fluctuations do not "average out" as the number of constituents becomes large. This is naturally a vast research programme, with a scope far exceeding what is achievable while preparing a doctoral thesis, and the work I present here is only a small attempt at studying some of the clues to solving this puzzle.

I have thus chosen to begin my dissertation with the study of power-law distributions. These distributions are ubiquitous throughout socio-economic data, and appear even in the frequency distribution of words in natural language. The first part of this dissertation studies these distributions, and attempts to give a comprehensive introduction to them by showing where they appear, the consequences they entail and, finally, a

survey of models that explain their origin. In particular, these distributions lead naturally to the breaking down of the central limit theorem, and so to anomalous statistical phenomena.

This is not however the whole picture: it is also necessary to understand how the different constituents of an economy are intertwined in the production network. The second part of my dissertation studies production networks, arguing in particular that the economy may *self-organize* into a state that is susceptible to amplify small shocks into large crises. This adds in the idea that fluctuations may be in part of *endogenous* origin, meaning that they can arise from the characteristics of the economy and because of how the different actors interact. This section also sets the stage for later work aiming to establish more precise models of the economy; such models should not only be able to qualitatively explain some features of the economy, but should also have the ultimate goal of being able to identify mechanisms in real data.

This necessity of understanding fine-grained economic data motivates the third part of my thesis, where I've studied firm-level data. In the corresponding Chapter, I have studied the statistical properties of firm growth rates and the dynamics of fundamental variables that describe them, such as their sales, and how they are different to those of "market-related" variables like their market capitalization. This constitutes the core of my empirical work.

Finally the last part deals with models that, in a way, provide alternatives to the idea of the rational representative agent that is often found in economic theory. These models show the surprising effects that emerge from taking into account interactions between agents and from going beyond the view that agents maximize a static utility function to make their choices.

Hoping to reflect the approach I have tried to have in these past 3 years, I have tried to make a good balance between data, analytical computations and numerical simulations in the writing of this thesis. I believe indeed that good science needs all three to attain a good understanding of observed phenomena. To me it is clear that empirical data is pre-eminent to everything, and that models should first and foremost try to explain whatever we see in it. Analytical derivations have the convenience of allowing one to shed light on certain features of the data, or of allowing to make logical deductions that go beyond those that mere language lets us do, but one should not let convenient axioms replace data as the grounding of such reasonings. Simulations in turn permit to generate *synthetic* data, allowing also for what I see as true experimental science, and to study problems that are much too complicated for analytical treatment, even if sometimes their inspection can lead to certain simplifications that lead to a simpler, solvable analytical model.[3]

3 The reader should also know that the data used for the plots, except confidential firm-level data, as well as all the scripts I wrote for the simulations in presented in this manuscript, are available upon reasonable request.

With this in mind, I have purposefully tried to write as to let all equations "blend in" with the rest of the text, have avoided to give analytical derivations any kind of distinction above data and simulations, and have also sought to give intuitive explanations to the phenomena and features I've studied whenever possible. My wish is above all that the reader finds a clear argumentation across this work, with the appendices in the last pages providing most of the mathematical justification behind it.

Part I

POWER LAWS

The opening Chapter of this thesis discusses power-law distributions. The reader will discover the stunning regularity that seems to pervade the statistical description of the frequency of words in natural language, the size of cities in a given country or the distribution of wealth among individuals, among many other quantities.

The reader will also become familiar with the practical implications of such distributions, both in terms of modelling and in a broader perspective.

Finally, the reader will see that they may often be explained by very simple models, and that the nature of power-laws gives often strong hints about the nature of the underlying generative processes, showing their importance in constraining the possible models to explain collective phenomena.

1

INTRODUCTION

A striking fact in socio-economic data is the prevalence of so-called power laws. In general, when thinking of a power-law linking two quantities x and y, one has in mind a relation of the type $y \propto x^{\alpha}$ where α is the exponent of the power-law, holding over a large array of values for x and y. Some relationships of this kind may seem quite trivial, as the one linking, for example, the side a of a cube with its volume V, as $V \propto a^{d}$, with $d = 3$. Importantly, this scaling relationship still holds if you consider instead the volume of a sphere of radius a or of any other regular three-dimensional object, for which you will always find the value $d = 3$.

Although in the case cited above there are obvious physical reasons for this type of scaling, it is startling to note that similar relations hold over other quantities without there being an evident cause. An eloquent example of this was found by G.K. Zipf [1] in what is now known as a Zipf law. Taking first James Joyce's 1916 modernist novel *Ulysses*, he showed that a scaling law emerged for the frequency f_k, or number of appearances, of the k-th most common word in the novel, given by

$$f_k \propto k^{-\alpha}, \quad \text{with } \alpha \approx 1. \tag{1.1}$$

Furthermore, he found that his "law" held over many different texts, roughly indistinctly of the language in which they were written.[2] This finding, which I've reproduced in Figure 1.1 along with other books, laid the ground for what has been called statistical or quantitative linguistics [73], in the hope of elucidating this and other regularities that seem to emerge in natural language.

Another, seemingly unrelated, "law" was also found in 1886 by Vilfredo Pareto when studying the distribution of wealth and incomes in Europe during the second half of the 19$^{\text{th}}$ century [208]. By taking a certain income level x, he counted the number of individuals $N_{>}(x)$ reporting an income larger than x and found that it followed the scaling relation

$$N_{>}(x) \propto x^{-\mu}, \quad \text{with } \mu \approx 1. \tag{1.2}$$

As with Zipf's finding, this relation held over a wide array of countries with different socio-economic characteristics. Although the exponent μ fluctuates somewhat between countries and in time, it is consistently found

1 Although Mandelbrot argues it was found earlier by Estoup [176, 177].

2 There are, naturally, variations to this. In particular, the exponent α fluctuates from language to language and from individual to individual. Mandelbrot noted that, as a child grows, the Zipf exponent describing his vocabulary decreases from roughly $\alpha \approx 1.6$ to $\alpha \approx 1.15$, or nearly 1 if said child happens to become James Joyce.

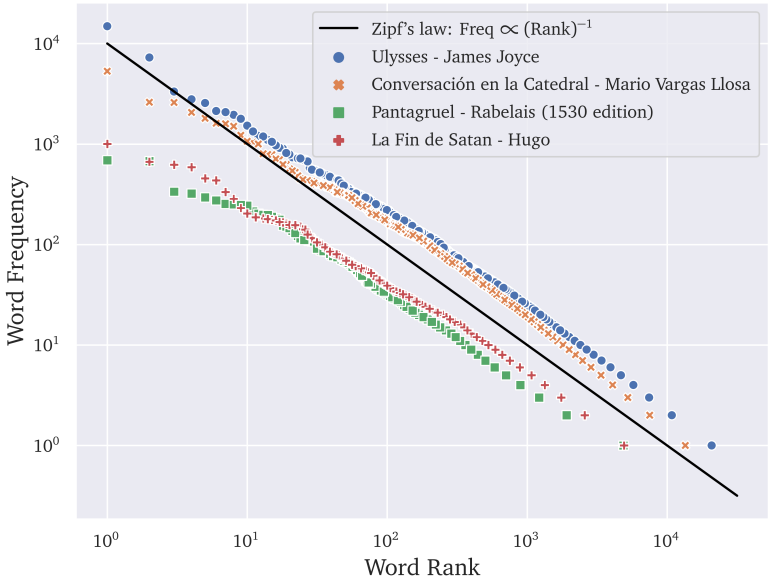


Figure 1.1: Rank-frequency plot of different texts. This suggests that the frequency f_k of the k-th most frequent word in a given text follows the scaling relation $f_k \propto k^{-\alpha}$ with $\alpha \approx 1$. Note the similarities between all curves, despite having been computed over texts with wildly different characteristics, from two somewhat modernist novels (Joyce and Vargas Llosa), to a novel written in 16$^{\text{th}}$-century French and a lengthy 19$^{\text{th}}$-century poem written entirely in French alexandrins.

to be close to 1. Note also that the exponent describing the distribution of income is consistently found to be larger than that describing wealth [267], a finding with important socio-economic consequences.[3] Using data from Pareto's book [208], I have reproduced his findings in Figure 1.2 for England and Ireland. As stressed above, despite the many differences between these two countries at the time, the relation in Eq. (1.2) holds well.

An objection may be raised, by saying that it is not obvious that the two relations in Eqs. (1.2) and (1.1), in addition to describing the behaviour of two completely different quantities, have anything to do with each other. Yet if, picking the example of income for definiteness, one ranks all individuals by income and defines $X_{(k)}$ as the income of the k-th richest

3 Indeed, as the reader shall see later, this means that wealth is more unequally distributed than income. The consequences of this are much beyond the scope of the present work, but are well summarized by what Piketty calls the *Rastignac* dilemma. Eugène de Rastignac is a character appearing in Balzac's *La Comédie Humaine*, a series of books set in 19$^{\text{th}}$ century France, whom the machiavelic Vautrin convinces to quit studying and marry a wealthy heiress instead, as acquiring a fortune allows for a much better existence than that afforded by the meagre income of a lawyer, however brilliant he may be. From a literary perspective, Balzac's *Le Père Goriot* and Zola's *L'Argent* give interesting views to the tension between income from work and wealth. For more information on this from an academic standpoint, the interested reader can look into Refs. [214, 215, 282].

Figure 1.2: The number of individuals $N_>(X)$ reporting an income higher than X in late 19$^{\text{th}}$ century England and Ireland. Both countries roughly satisfy the scaling relationship given in Eq. (1.2).

individual in the sample, you immediately have that $N_>(X_{(k)}) = k$. From Eq. (1.2) one then immediately has the relation

$$X_{(k)} \propto k^{-1/\mu}, \tag{1.3}$$

showing then that the representations of Figures 1.1 and 1.2, as well as the corresponding Eqs. (1.1) and (1.2), are in fact equivalent by setting $\alpha = 1/\mu$.

This equivalence, and another example for the list of socio-economic quantities that seem to follow power-laws, is illustrated in Figure 1.3, where there seems to be a Pareto/Zipf scaling relation for the population of US cities. One sees again that the exponent μ is close to 1.

My main interest will therefore be on *power-law distributed quantities*, that is those quantities for which the statistical distribution follows the scaling relation described above.

The number of quantities that follow similar scaling relations (see e.g. [88]) is impressive. It's the case of the degree distribution of protein interaction networks in biology, the degree distribution between different nodes in the internet, the number of telephone calls received by people in the United States, the intensity of wars as measured by the number of battle casualties, the number of species per genus in mammals, or even the amount of citations received by scientific papers. Interesting reviews can be found for example in [119, 229].

The pervasiveness of power-laws in socio-economic phenomena seems therefore so strong that one ends up wondering if there is anything actually

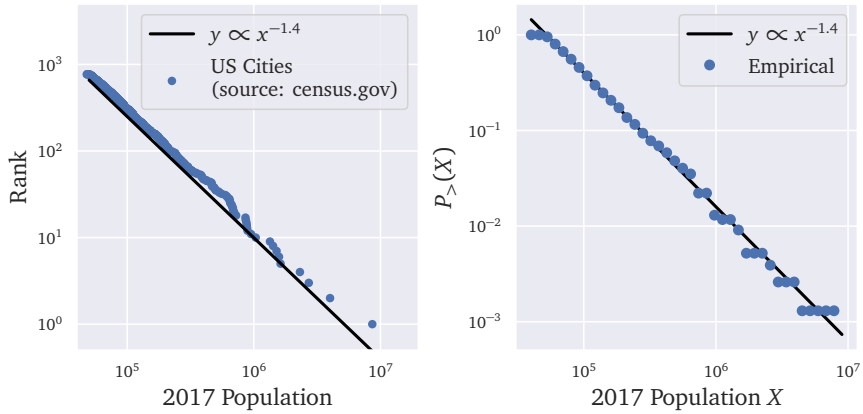


Figure 1.3: Size-rank and size-cumulative plot for US Cities population in 2017. Note the scaling relation as in Eq. (1.3) and Eq. (1.2) with $\mu \approx 1.4$.

special about them. The goal of the present Chapter is to present an overview of power-laws as a motivation for the modelling done in the rest of this thesis, and to make a strong case that there is an interest in understanding them, the ways they emerge and the consequences they entail.

The goal is threefold, and I shall seek to emphasize that: (i) quantities distributed according to power laws lead to counter-intuitive statistical phenomena, (ii) there are important socio-economic implications for these phenomena, and lastly (iii) the different statistical mechanisms that explain the presence of power-laws in many of the quantities described above and in the literature have overarching principles. This last point is, in my view, the most important in this chapter, and sets the epistemic tone for the rest of this thesis, related to the leitmotiv of statistical mechanics of finding statistical laws that encompass the behaviour of seemingly different collective systems.

2

STATISTICAL PROPERTIES OF POWER-LAWS

This section will discuss the various statistical properties of power-laws, as well as the immediate consequences they have on different topics such as risk control and wealth inequality.

I will first give a broad definition of a power-law distribution, before delving into the different properties of these distributions.

2.1 SOME DEFINITIONS

Confronted with a set of observations $(X_1,\dots,X_N)$ of, say, city sizes or the incomes, one says that X is power-law distributed or that X is a power-law if the relation (1.2) holds at least for $X > X^+$ for a certain value $X^+ \gg 1$. When the number of observations N is large, it is easier to take the continuous limit and consider the quantity $P(x)$, defined by

$$P(x) = \lim_{N\to\infty} \lim_{\Delta x\to 0} \frac{N_>(x+\Delta x) - N_>(x)}{N}, \tag{2.1}$$

where N is the total number of individuals in the sample. One can then also define the complementary cumulative distribution function (complementary cdf), or survival function, as:

$$P_>(x) := \int_x^\infty \mathrm{d}s\ P(s), \tag{2.2}$$

which one sees immediately satisfies also $P_>(x) \approx N_>(x)/N$. This function is related to the cumulative distribution $P_<(x)$ function by the relation $P_>(x) = 1 - P_<(x)$.

Thus, $P(x)$ becomes the probability density associated to a random variable X. As a consequence of Eq. (1.2), one sees that one should have $P(x) \propto x^{-1-\mu}$ or, equivalently, $P_>(x) \propto x^{-\mu}$ over a range $x > X^+$. If it holds over the whole range of observations $[x_{\min}; +\infty[$, then one has a pure power-law distribution, also called a Pareto distribution, i. e.

$$P(x) = \frac{\mu x_{\min}^\mu}{x^{1+\mu}}. \tag{2.3}$$

The case of a pure power-law distribution seems rather restrictive to circumscribe all the distributions which are generally called power-laws. In particular, a pure power-law would imply a perfectly straight line in Figures 1.1,1.2 and 1.3. For this reason, the definition I adopt of a power-law

distribution is a distribution such that $P(x|x > X^+)$, i.e. the probability density of S conditional on being larger than X^+, is reasonably well approximated by a Pareto density as in Eq. (2.3) for some μ, which is called the exponent of the power-law, and with $x_{\min} = X^+$. This highlights that the power-law behaviour is something that is defined in the *tails* of the distribution, and is equivalent to say that X is power-law distributed if the associated probability density function verifies

$$P(x) \underset{x \gg 1}{\sim} x^{-1-\mu}. \tag{2.4}$$

Naturally, although I have introduced definitions for a non-negative random variable $X > 0$, one can extend everything real-valued random variable X, with possibly two different exponents μ_+ and μ_- describing the right and left tails, as a variable whose density satisfies

$$P(x) \approx \begin{cases} A_+ x^{-1-\mu_+} & \text{for} \quad x \gg 1 \\ A_- x^{-1-\mu_-} & \text{for} \quad x \ll -1 \end{cases}, \tag{2.5}$$

with $A_\pm$ some positive constants.

This now allows to explore the main statistical properties of power-laws.

2.2 MOMENTS AND EXTREME VALUES

The first consequence of the definition Eq. (2.4) is that the moments of the variable X are not always defined. If one wishes to compute a moment $\mathbb{E}[X^m] = \int \mathrm{d}x \; P(x) x^m$ then it is immediate to see that the integrand behaves as $x^{m-1-\mu}$ for large x, and therefore that the integral converges only for $m > \mu$.[1] As a consequence, the standard deviation or variance are *not defined* for a power-law distribution with $\mu < 2$, nor is the mean when $\mu < 1$.

Naturally, a real dataset has a finite number of observations N, and so the empirical moments are always finite. The divergence of the moments should then be interpreted in this setting. Having a set of N observations $(X_1, \dots X_N)$, one would expect, because of the law of large numbers, the convergence:

$$\frac{1}{N} \sum_{i=1}^{N} X_i^m \underset{N \to \infty}{\longrightarrow} \mathbb{E}[X^m]. \tag{2.6}$$

However, the statement above implies that the sum on the left hand side diverges as $N \to \infty$ when $m \leq \mu$. By taking the case $\mu < 1$ and looking at the behaviour of the empirical mean one can in fact gain further intuition by

1 In fact this can be used as an alternative definition for a power-law distribution: one defines μ as the infimum of $\{m | \mathbb{E}[S^m] < \infty\}$, the distribution is a power-law if $\mu < \infty$ and one has directly the definition of μ. Note that this definition also says that a distribution with a density that behaves asymptotically as $P(x) \approx \ln(x)^\alpha x^{-1-\mu}$ is a power-law distribution. In practice, it is very difficult to distinguish this logarithmic behaviour when doing an empirical analysis.

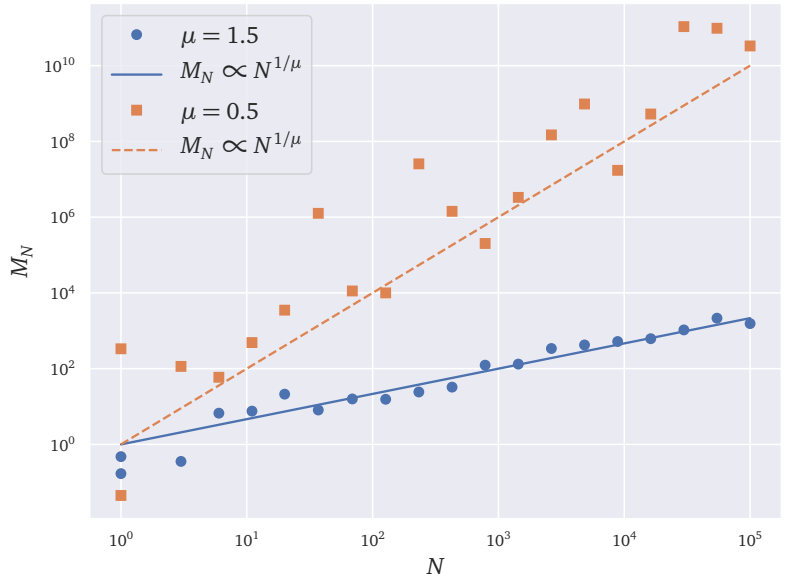


Figure 2.1: The maximum of N pure power-laws vs. N for two different exponents. Note that the behaviour described by Eq. (2.7) holds.

noting that the largest value, denoted M_N, should satisfy $P_>(M_N) \approx 1/N$. Using Eq. (2.2) one then immediately has

$$M_N \sim N^{\frac{1}{\mu}}, \tag{2.7}$$

which means that the ratio M_N/N behaves as $N^{\frac{1-\mu}{\mu}}$ and therefore diverges as $N \to \infty$. The left-hand side of Eq. (2.6) is then necessarily divergent, but this result allows us to make a stronger claim: the sum of the X_i variables consists of N terms, the largest of which scales as $N^{\frac{1}{\mu}}$ and one can therefore note that the largest observation makes up for a considerable fraction of the aggregate of all observations. I shall come back to this when considering the repercussions of power-laws on inequality. The same reasoning holds when considering moments such as X^m by noting that the variable $Y = X^m$ is a power-law with a density $P(y) \propto y^{-1-\frac{\mu}{m}}$.

This should be contrasted with e. g. an exponential distribution. Take a random variable X with density $P(x) = e^{-x}$, and so with a complementary cdf given by $P_>(x) = e^{-x}$. Using the same reasoning as above one finds that the largest observation N grows rather slowly, as $M_N \sim \ln(N)$.[2] Notably, $M_N/N \to 0$ in the large N limit in this case, implying that the largest observation makes only for a vanishing fraction of the aggregate in the large N limit.

2 Note that the maximum of N gaussians also grows slowly with N, as $\sqrt{2\ln N}$.

2.2.1 *Maxima of random variables*

One can in fact make precise statements regarding the maximum of power-law distributed variables. Consider N observations from a distribution with a tail behaving as

$$P_>(x) \approx \frac{A^\mu}{x^\mu}, \tag{2.8}$$

with the largest observation M_N. Then one can define a rescaled variable $u = \frac{M_N}{AN^{1/\mu}}$, so defined to be of order 1, and show using extreme-value theory that the distribution of u converges as $N \to \infty$ to the Fréchet distribution, defined as

$$P(u) = \frac{\mu}{u^{1+\mu}} e^{-u^{-\mu}}, \tag{2.9}$$

which one can note is also asymptotically a power-law distribution (the exponential goes to 1 for large u) with the *same* exponent μ. To understand how this comes about, one should note that the cumulative density function of M_n verifies

$$\mathbb{P}(M_N < AN^{1/\mu}x) = \left(1 - P_>(AN^{1/\mu}x)\right)^N \approx \left(1 - \frac{x^{-\mu}}{N}\right)^N, \tag{2.10}$$

which means that in the large N limit the cdf of u reads $P_<(u) \approx e^{-u^{-\mu}}$. One finds then Eq. (2.9) by differentiation. This in fact suggests the scale-invariant nature of power-laws, more comments below.

Note also that these statistical properties of extreme values are also different when one considers random variables with a thin tail, i. e. those whose density goes to 0 faster than a power-law for large x. Following [65], take for example a random variable with a density

$$P(x) \approx \frac{A}{x^\alpha} \exp\left(-Bx^\delta\right), \quad B, \delta > 0, \tag{2.11}$$

which one can show has a complementary cdf given for large x by

$$P_>(x) \approx \frac{\delta A}{B} x^{1-\delta} \exp\left(-Bx^\delta\right). \tag{2.12}$$

This means that to leading order the scaling of the maximum is given by $M_N \sim \left(\frac{\ln(N)}{B}\right)^{1/\delta}$. Introducing then the variable u such that

$$M_N = \left(\frac{\ln(N)}{B}\right)^{1/\delta} + \left(\frac{\ln(N)}{B}\right)^{-1/\delta} \frac{u}{B\delta}, \tag{2.13}$$

then one has that the distribution of u is given by the Gumbel distribution,

$$P(u) = \exp(-u - \exp(-u)), \tag{2.14}$$

meaning that in this case the maximum has *exponential* tails.

2.3 SUMS OF VARIABLES

When considering sums of the kind seen in Eq. (2.6), such as

$$S_N = \sum_{i=1}^{N} X_i, \tag{2.15}$$

one is tempted to invoke the standard version of the central limit theorem, and write for large N that

$$S_N = N\mathbb{E}[X] + \sqrt{N\mathbb{V}[X]}\xi, \tag{2.16}$$

where ξ is a centred gaussian random variable with unit variance. This depends necessarily on the existence of the moments $\mathbb{E}[X]$ and $\mathbb{E}\left[X^2\right]$, and entails that the "standard" central limit theorem (CLT) breaks down for power-laws with $\mu \leq 2$.

One can nonetheless extend the central limit theorem using so-called Lévy-stable distributions. Consider a power-law distributed variable X with $P(x) \approx Ax^{-1-\mu}$ for large x, then the generalized central limit theorem reads

$$S_N = \begin{cases} N\mathbb{E}[X] + (AN)^{1/2}\,\xi_G & \text{for} \quad \mu > 2 \\ N\mathbb{E}[X] + (AN)^{1/\mu}\,\xi_L & \text{for} \quad 1 < \mu \leq 2\,, \\ (AN)^{1/\mu}\,\xi_L & \text{for} \quad \mu < 1 \end{cases} \tag{2.17}$$

where ξ_G is a gaussian random variable and ξ_L a variable drawn from a Lévy-stable distribution with parameter μ. The reason behind the difference in the scalings in $\sqrt{N}$ for Eq. (2.16) and in $N^{1/\mu}$ can be understood by reasoning on the typical fluctuations about the mean of S_N: when the variance is defined, fluctuations behave as in the gaussian case and scale as $\sqrt{N}$; in the power-law case however these fluctuations can be imputed to the maximum of the X_i variables, which scales as $N^{1/\mu}$.

This rationale also helps us understand the behaviour of the Lévy distributions, which have power-law tails themselves, as their distribution for large ξ is given by:

$$L(\xi) \approx \frac{\mu A}{|\xi|^{1+\mu}}. \tag{2.18}$$

Unsurprisingly, because of what I argued above for the Fréchet distribution for the maximum of N power-law distributed variables as a power-law itself, the Lévy-stable distribution also has power-law tails with the same exponent.

A sketch of the proof for the generalized central limit theorem is given in Appendix A, Section A.1.

When $\mu > 2$, you could ask, because of the previous arguments on the maxima of N random power-laws, how the stability of the maximum is compatible with the gaussian distribution. In other words, one expects that

for $\mu > 2$ the largest term in S_N/N should scale as $N^{1/\mu-1}$, and therefore go to 0 much slower than in the gaussian case where one would expect $M_N/N \sim \sqrt{\ln(N)}/N$. The answer to this lies in the interpretation of the central limit theorem: one sees essentially that I used convergence arguments for the characteristic function at small t, but did not show at which speed the corrective terms vanished as $N \to \infty$.

The distribution of S_N doesn't in fact converge *uniformly* to a gaussian distribution, and that approximation is only valid in a central region which one can show grows, in the case of power-laws, as $\sqrt{N\ln(N)}$. The argument for this is simple, if there is a power-law decay then one expects that for large S the probability amplitude of S_N should scale as $\sim NS^{-1-\mu}$, while for the gaussian region it should scale as $\exp\left(-S^2/2N\sigma^2\right)$. Equating these two amplitudes to find the boundary of the central region S^* leads to $S^* \sim \sqrt{N\ln(N)}$.

2.3.1 *An interesting property: condensation*

It may sometimes be of interest to compute sums of moments of a power-law X, or equivalently to compute q-norms by looking at objects such as

$$S_{N,q} = \left(\sum_{i=1}^{N} X_i^q\right)^{1/q}. \tag{2.19}$$

By conducting the same type of reasoning as above, one sees that the largest value in the sum inside the parentheses should scale as $N^{q/\mu}$, and that it is consequently dominated by the largest value for $q > \mu$. In this case, the norm condenses on the largest value, i. e.

$$S_{N,q} \approx X_{\max}, \quad \text{for } q > \mu, \tag{2.20}$$

a situation that is analogue to a phase transition in statistical physics, where q plays the role of an inverse temperature and $S_{N,q}$ is akin to a partition function.

An interesting consequence comes when studying the so-called herfindahl index used in economics [137], which is in fact also linked to the Inverse Participation Ratio in condensed matter physics [270, 41], which will come up again in Chapters 5, 6 and 8. The herfindahl index associated to variables $X_1, \ldots, X_N$ is

$$H = \frac{\sum_{i=1}^{N} X_i^2}{\left(\sum_{i=1}^{N} X_i\right)^2} = \sum_{i=1}^{N}\left(\frac{X_i}{\sum_{j=1}^{N} X_j}\right)^2, \tag{2.21}$$

which, through the same argument as the one used above, should be of order 1 for $\mu \lessapprox 2$.

The herfindahl index is in fact used as a measure for concentration: if all the values X_i are evenly distributed then $X_i \sim 1/N$ and so $H \approx 1/N$, while

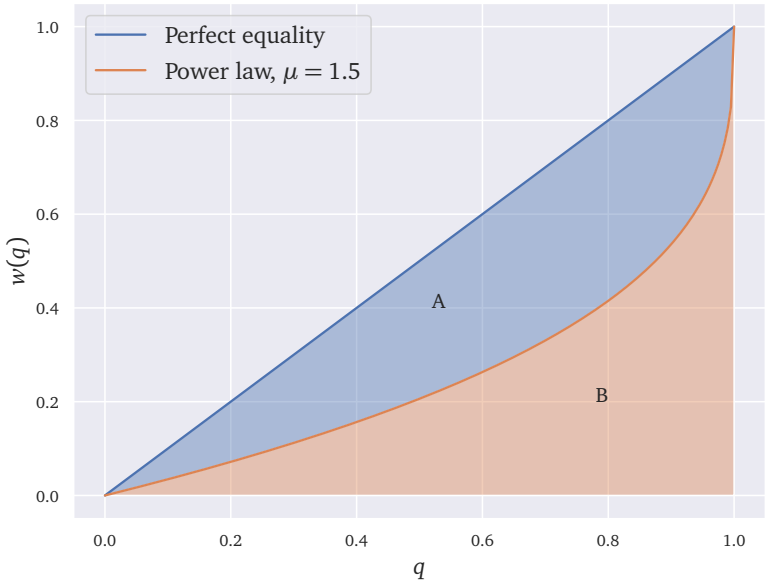


Figure 2.2: An illustration of the Gini index for a power-law distribution for wealth with $\mu = 1.5$. The Gini index is defined as the ratio of the areas $G = A/(A+B)$. The x axis describes the quantile q while the y axis gives the fraction of the total wealth $w(q)$ owned by that quantile. The top curve corresponds to $w(q) = q$, a situation with perfect equality, while the bottom curve corresponds to a situation where the wealth is power-law distributed.

for a perfectly concentrated sample you should have $X_i / \sum X_j \approx 0$ for all i, except for a certain value $X_{\max}$ for which this should equal 1. The quantity $1/H$ can therefore be used as a proxy to see how many observations X_i effectively contribute to the aggregate $\sum X_i$.

2.4 INEQUALITY AND SCALE INVARIANCE

For the following section, I restrict the analysis for simplicity to the case where one has X_i variables that follow a pure power-law as defined in Eq. (2.3). For clarity of example, I shall talk about X_i as describing the wealth of an individual.

Within this setting, the total wealth in a society of N individuals is given by $W_N = \sum_{i=1}^{N} X_i$. The question to ask is then how evenly the total wealth is distributed among the N individuals. As said above, the question is trivial when $\mu < 1$, as the wealth of the richest individual is of the order of $N^{1/\mu} \gg N$, which entails that the total wealth is possessed by at best a handful of individuals.

In the case $\mu \geq 1$, one can compute the total wealth in the large N limit through the law of large numbers, which should equal $N\mathbb{E}[X]$. Choosing

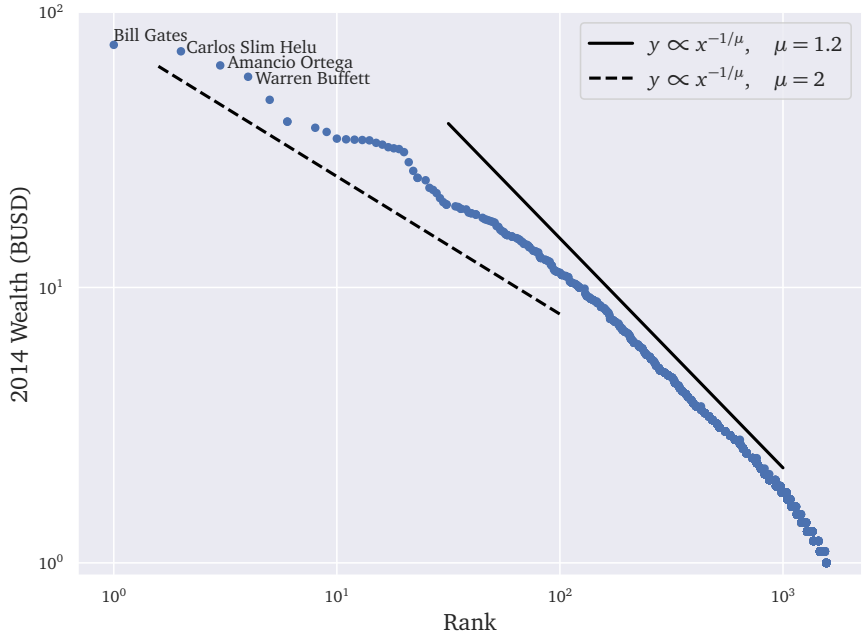


Figure 2.3: Zipf plot describing billionaires and their wealth in 2014. Data obtained from [76].

$p(x) = \frac{\mu}{x^{1+\mu}}$ (i.e. taking $x_{\min} = 1$ in Eq. (2.3), or equivalently counting wealth in units of $x_{\min}$), one finds

$$W_N \approx N \frac{\mu}{1-\mu}. \tag{2.22}$$

One can now ask what the wealth of the q-th quantile is, having defined the quantile as the fraction $q \in [0;1]$ of poorest individuals. This wealth level x_q satisfies $P_<(x_q) = \int_1^{x_q} \mathrm{d}x\ p(x) = q$, which can be inverted to find $x_q = (1-q)^{-1/\mu}$. The share of the wealth owned $w(q)$ by the q poorest individuals is then

$$w(q) := \frac{N \int_1^{x_q} \mathrm{d}x\ p(x)x}{W_N} = 1-(1-q)^{\frac{\mu-1}{\mu}}, \tag{2.23}$$

where one sees that the perfect equality $w(q) = q$ is reached in the $\mu \to \infty$ limit, while $w(q)$ goes to 0 everywhere except at $q = 1$ in the limit $\mu \to 1$, corresponding to the wealth condensation described above, in a *winner takes all* setting.

With this definition one can then define the well-known Gini index, used in the economics literature to quantify the level of inequality within a country, as the area between the curve $(q, w(q))$, also known as the Lorenz Curve, and the curve (q, q). A direct computation gives the Gini index as a function of the tail exponent μ,

$$G(\mu) = \frac{2\mu}{2\mu-1} - 1, \tag{2.24}$$

and goes to 0 as $\mu \to \infty$, as one expects. An illustration of the definition of the Gini index is shown in Figure 2.2.

An interesting setting is also popularly known as the "80/20 rule", stating, for example, that the top 20% richest individuals possess 80% of the wealth, or that the 20% most common words in a text make up for 80% of its physical content. This is equivalent to saying $w(0.8) = 0.2$, and one can easily solve to find the relevant tail exponent $\mu = \frac{\ln(5)}{\ln(4)} \approx 1.16$, very close to the exponents found in empirical data. As an illustration, I have plotted in Figure 2.3 a Zipf plot describing billionaires and their wealth in 2014, although it should be admitted that wealth data is notoriously hard to come by and that there may be gross misestimates.

Another interesting consequence is that of scale invariance. If one considers the ratio between the wealth share owned by the $1-q$ richest individuals it is given by $1-w(1-q) = q^{\frac{\mu-1}{\mu}}$, while on the other hand the $1-aq$, with $a < 1$, richest individuals own a share of the wealth equal to $a^{\frac{\mu-1}{\mu}} q^{\frac{\mu-1}{\mu}}$. The ratio between the two quantities depends only on a, and not on q. An example of this is that, say, the ratio between the share owned by the top 10% and the total wealth is the same as the ratio of the share of the wealth of the top 1% divided by the share of the top 10%.

The scale invariance is in fact encoded in the very definition of a pure power-law. Consider indeed a variable $X > 1$ distributed as $p(x) = \frac{\mu}{x^{1+\mu}}$. Then the distribution of X conditional on being larger than x^* is quite simply given by

$$\frac{p(x)}{\int_{x^*}^{\infty} \mathrm{d}y \; \mu y^{-1-\mu}} = \frac{\mu (x^*)^{\mu}}{x^{1+\mu}} \tag{2.25}$$

for $x > x^*$, and 0 otherwise. This is therefore a power-law with the same exponent μ of the general form defined in Eq. (2.3), and the conditioning only had the effect of introducing a different lower cut-off, implying that this also holds for distributions that are only asymptotic power-laws. Another interesting property is that the variable $Y = aX$ for some value of a is itself also power-law distributed with the same exponent μ as that of X. This means that the exponent doesn't, for example, depend on the unit of measurement of X. Note that this isn't the case for exponential variables: if X is exponentially distributed, then the distribution of X conditional on being larger than x^* has the same exponential tails, but if one takes a rescaled variable $Y = aX$ then the distribution of Y is still exponential, but the exponential decay is multiplied by a factor $1/a$.

2.5 SUMMARY AND REMARKS

The present section has attempted to show that power-laws tend to emerge in a lot of socio-economic phenomena. These are not mere statistical curiosities, but have deep implications: a quantity that is power-law distributed behaves in a significantly distinct manner than if it were described by a gaussian distribution. As a clear example one may think about the different

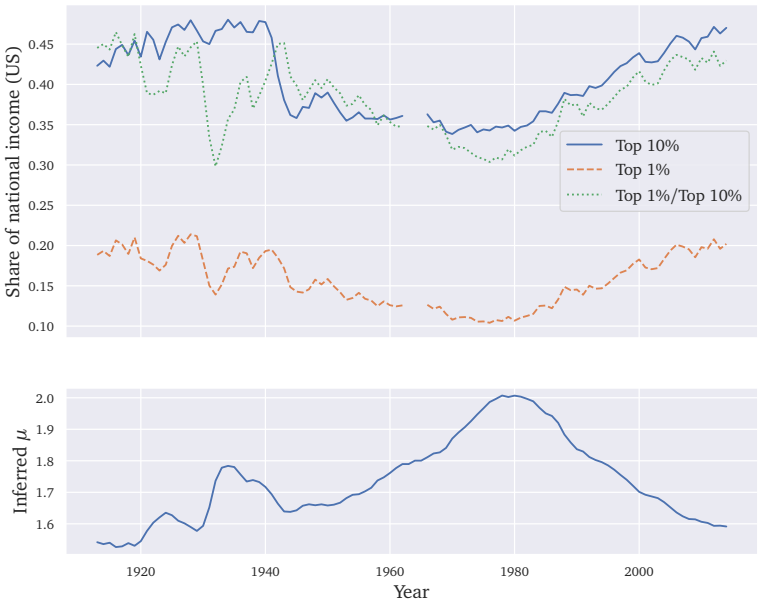


Figure 2.4: Income before tax distribution data for the United States from 1913 to 2014. I show the share attributed to the 10% and 1% highest income individuals as well as the ratio between these. Note that the ratio closely follows the curve for the share of the 10%, as one would expect in a power-law distribution. One can then infer the tail exponent from this ratio by equating it to $(0.1)^{\frac{\mu-1}{\mu}}$. The inferred exponent fluctuates between $\mu \approx 1.5$ to $\mu \approx 2$ and seems somewhat larger than that of Fig. 1.2, and one sees the pattern of falling inequalities in the post Second-World War period followed by a subsequent increase since the 1980s decried by Piketty in [214] and [215]. Data sourced from [216].

ways that height and weight are distributed: one can reasonably say that the average height of a human adult is of $\approx$ 165cm, and that most people fall within, say, $\pm$30cm of that mark. You would reasonably expect that the height difference between the tallest person on Earth and the 500 tallest is of the order of a few centimetres only, and it would be unthinkable to imagine someone that measures 500 times the average height!

In contrast, the French magazine *Challenges*' 2017 ranking of French fortunes estimated the fortune of Bernard Arnault at around 46, 900 million euros and placed him first, while the 500th fortune, Jean-Baptiste Rudelle's, was estimated at 120 million euros. The ratio between the two is of about $1/500$, as expected if one assumes that wealth follows a Zipf law but completely different to what one would see if wealth were distributed according to a gaussian distribution.

The implications of this are deep, in particular regarding the modelling of economic agents. With such a stark heterogeneity between the income and

wealth of individuals one can only think of the proverbial "representative agent" as a theoretical construct that is far removed from reality.

This is not the only problem stemming from power-laws that affects modelling economic or social systems. Indeed, a great many models in economics and finance conceptualize the systems they study as being semi-deterministic, with random fluctuations pushing economic variables about their "equilibrium" values or an asset's price about its "fundamental" value, whatever those concepts may mean. These fluctuations are then thought of as being gaussian, regrettably neglecting the fat-tailed nature of the system and sometimes leading to an underestimation of risk.

As I have argued, the sum of N power-law variables with $\mu > 2$ can only be considered to be gaussian-distributed in a region that grows very slowly, as $\sqrt{N \ln(N)}$, with the region beyond this value still described by a power-law. When the exponent μ is smaller than 2 this gaussian approximation breaks down completely, and one no longer has well-behaved gaussian fluctuations. As the reader shall see, the behaviour of power-laws has strong consequences for macroeconomic fluctuations under what is best known as *granularity*.

To the modeller, the ubiquitousness of power-laws should act as a permanent warning to avoid the overuse of gaussian intuition. Yet this is not the only lesson to be learnt from them. As I will show, these distributions can also be used both as guides into the basic mechanisms describing a system and as tools to discriminate models: obtaining gaussian fluctuations from a stochastic model is often not a difficult task, but this isn't the case for power-law fluctuations, and one can confront the nature of the distributions obtained from a model with available data, as a discriminatory criterion to keep it or discard it.

3

MECHANISMS FOR POWER-LAWS

In this chapter, I will try to illustrate different models that lead to power-law distributions. Because of their historical significance, I will first study models that attempt to explain the emergence of the Zipf law for word frequencies.

Most of these explanations lead naturally to models of *stochastic multiplicative growth*, which under very mild assumptions lead to power-laws. For these models, the exponent μ depends on different parameters describing the growth of entities, as does the rate of convergence of the stochastic process to its stationary distribution. The Zipf case $\mu \approx 1$ often appears as a limiting case.

Finally, I will explore a different scenario for the generation of power-law distributions, showing that they emerge in large systems close to *criticality*, a notion that I will define below. Importantly, in this last case the exponent of the distribution is no longer related to microscopic parameters, but falls instead into *universality classes* that describe the coarse-grained nature of the system under study, just as the exponent 3 describes the scaling relationship between the characteristic length of a 3-dimensional object and its volume, as explained in the introduction with the examples of a cube or a sphere.

3.1 POWER-LAW MODELS FOR NATURAL LANGUAGE

Because of the stunning regularity with which the Zipf law appears in language, many models where conceived to explain this statistical fact. They strive to explain power-laws with exponent $\mu \approx 1$ using mechanisms that rely on heuristics based on natural language. Although this may seem irrelevant to the study of scaling relations and power-laws in, e. g. , economics, the reader shall see that these models may be transposed onto different systems, and that they are therefore relevant in the study of generic mechanisms leading to power-laws.

3.1.1 *Monkey typists*

Perhaps surprisingly, one of the simplest arguments to explain the pervasiveness of the Zipf law in natural language, as illustrated in Fig. 1.1, is that it could be generated by monkeys typing randomly on a keyboard.

This was highlighted first by Miller [186], although more sophisticated versions have been recently studied [210, 50]. His original model runs as follows.

Consider a keyboard with n letters, which I represent as L, and a space bar S. A monkey presses the space bar with a certain probability p, or any other letter L with probability $q = 1-p$. Any given word can be represented as $L_1, \ldots, L_i S$, where i is the length of the word, and so the probability of having a word of length i is given by

$$P_i = q^i p. \tag{3.1}$$

For any given word length i however, there are n^i different possibilities. If one considers a specific word $w = L_1, \ldots, L_i S$ of length i, then the expected frequency at which it should appear is

$$f_w = P_i / n^i = \left(\frac{p}{q}\right) e^{-i(\ln(N)-\ln(q))}, \tag{3.2}$$

showing, as visible already in Eq. (3.1), that the frequency of appearance of a word falls of exponentially with its length in this case.

Miller then says that all words of one letter have ranks k_w between 1 and n, words of two letters will be ranked between $n+1$ and $n+n^2$ and so on, with in general a word of i letters being ranked between $\frac{n(1-n^{i-1})}{(1-n)} + 1$ and $\frac{n(1-n^i)}{(1-n)}$. This allows him to replace the rank of any given word of length i by its "average rank", or the middle of the interval above, which leads to the equation

$$f_w = b\,(k_w + c)^{-\alpha}, \quad \alpha = 1 + \frac{\ln(1-p)}{\ln(n)}, \tag{3.3}$$

indicating an asymptotic power-law with an exponent $\mu = 1/\alpha$. Empirically, it is possible to test this by generating random strings of text from a certain alphabet, and one sees that the Zipf behaviour is well reproduced [166].

Naturally, one may object that this model isn't realistic: the most frequent word in English is the word "*the*", followed by "*of*" and "*and*", an observation that seems incompatible with the suggestion that the frequency of a word falls of exponentially with its length. Amendments can be attempted, as in [50], by stating that the n letters in the alphabet can be chosen different probabilities $q_1, \ldots, q_n$, yet this still feels somewhat artificial. Taking, for example, Mario Vargas Llosa's *Conversación en la Catedral*, which I used for Fig. 1.1, I've found that the most common word in the novel is "*que*" (that), despite the letter q not being very frequent in Spanish words, and that the fourth one is "*dijo*" ([s/he] said), something that isn't altogether unexpected in a novel bout a conversation between two persons, who themselves discuss other conversations and events.

In short, one may argue that this model doesn't take into account grammar or context. However, the most interesting result with it is that the exponent μ gets closer and closer to the "ideal" Zipf value of 1 as the probability of hitting the space bar $p \to 0$. Strikingly, the reader will see that this is similar to what happens in many other models.

3.1.2 *The Simon model*

The following model was proposed by Herbert Simon in [242], and puts forward a different mechanism of random text generation.

Imagine writing a text that so far has t words, some of them repeated instances of the same word, and that you are about to write another one down. The Simon model states that it will be a novel word with a certain probability α, or an instance of a previously existing word that will be uniformly chosen among the first t with probability $1-\alpha$. There is therefore a certain number of distinct words, each of them with a given number of appearances in the text. The model therefore stipulates that the number of appearances of each distinct word will grow proportionally to its current number of appearances, as words that have already been repeated a lot will be chosen more often. One can then write the evolution equations for $N(k,t)$, the number of words having appeared k times in a text of length t, as

$$N(k,t+1)-N(k,t)=\frac{1-\alpha}{t}\left[(k-1)N(k-1,t)-kN(k,t)\right]+\alpha\delta_{k,1}, \tag{3.4}$$

indicating that the difference comes either from appearing k times and not being repeated, or from appearing $k-1$ times and being repeated once.

The frequency of appearances of each word is then $p(k,t)=N(k,t)/(\alpha t)$, as there are on average αt distinct words as $t \gg 1$. I then take a continuous limit by writing $N(k,t+1)-N(k,t) \approx \frac{\partial N(k,t)}{\partial t} = \partial_t \left[N(k,t)\right]$, getting

$$\frac{\partial\left[t\ p(k,t)\right]}{\partial t}=\delta_{k,1}+(1-\alpha)\left[(k-1)p(k-1,t)-kp(k,t)\right]. \tag{3.5}$$

For large values of k, the second term on the right-hand side may be written as $-(1-\alpha)\partial_k\left[k\ p(k,t)\right]$. Assuming then that for $t\to\infty$ the distribution converges to a stationary state denoted $P(k)$, one finds that it solves the differential equation

$$p(k)=-(1-\alpha)\frac{\mathrm{d}\left[k\ p(k)\right]}{\mathrm{d}k}. \tag{3.6}$$

Surmising a solution of the form $p(k)\approx Ak^{-1-\mu}$ for large k therefore leads to

$$Ak^{-1-\mu}=A\mu(1-\alpha)k^{-1-\mu}, \tag{3.7}$$

and so

$$\mu=\frac{1}{1-\alpha}, \tag{3.8}$$

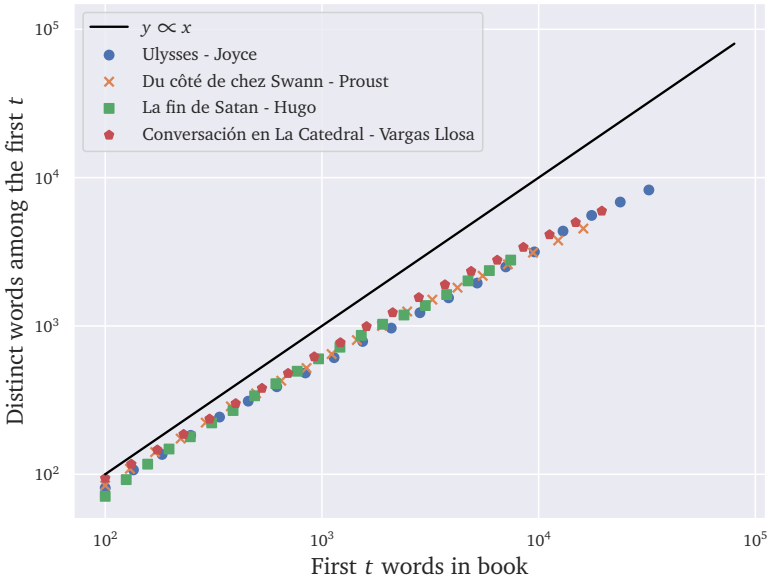


Figure 3.1: A test of one of the consequences of Simon's model. According to it, a text of length t (in words) should have αt distinct words. I have checked this with the texts used in Fig. 1.1, and see that the number of unique words among the first t scales as $\propto t^{\beta}$ with β slightly under 1. This is actually known as Heap's law, see e. g. [11], and it is consistently found that the number of distinct words in a text of length t scales as t^{β} with $\beta < 1$.

corresponding also to the exact result obtained by Simon, namely

$$p(k) = \mu \frac{\Gamma(1+\mu)}{\Gamma(1+k+\mu)} \underset{k \gg 1}{\propto} k^{-1-\mu}, \tag{3.9}$$

where $\Gamma(x)$ is the Gamma function.

For small α, one may write $\mu \approx 1 + \alpha$ to get a Zipf law in the limit $\alpha \to 0$, viz. the limit where no new families are added.

Interestingly, it is possible to compare the prediction of Simon's model, where the number of distinct words in a text of length t grows as αt, to what happens in the books from Fig. 1.1. The results can be seen on Fig. 3.1, where the linear scaling predicted by the model is not too far from what actually happens. This should again be linked to the result at the end of the previous Section 3.1.1.

The Simon model is in fact equivalent to the Yule model, which I will describe later on.

3.1.3 *Sample-space reducing models*

An interesting aspect from the previous Simon model is that, as context emerges in the generation of a text, words that have already appeared are

strongly favoured over new ones. A similar idea would be to say that the choice of the next word in the text depends strongly upon the last word. For example, imagining that one starts a sentence with the word "He" implies that the next word is likely going to be a verb, just as having the beginning of a sentence speak of a dog greatly constrains the possible words one may choose for its end.

This is essentially the idea behind sample-space reducing (SSR) processes, introduced by Corominas-Murtra et al. in [94] and [263, 138]. Starting from an initial "word" in a set Ω_i, one may pick the next word only from subsets Ω_j, such that $\Omega_j \subset \Omega_i$. Thus, the set of words from which one is sampling gets smaller and smaller. The simplest version of the model has one pick one of these subsets uniformly. It can actually be mapped into a continuous system, showing its link to multiplicative processes, an original contribution done in the paper, [190].[1]

The idea is in fact to map a set Ω_i to a set $(-\infty; E]$, with E an "energy" level. One may then pick the next word from a set $\Omega_j \subset \Omega_i$, meaning that the next choice amounts to picking $E' < E$, and so the problem is mapped to the dynamics of a system that always goes into lower and lower energy levels. The energy levels are therefore distributed with a distribution $\mathscr{P}(E)$, counting the number of available levels between E and $E + \mathrm{d}E$, with a normalization $\int \mathrm{d}E\, \mathscr{P}(E) = 1$. In this setting, $\mathscr{P}_<(E) = \int_{-\infty}^{E} \mathrm{d}E'\, \mathscr{P}(E')$ is proportional to the number of choices that are currently available, and as the dynamics progresses this quantity gets smaller and smaller, as the number of available choices reduces. Each word is therefore mapped to an energy level, with lower-energies corresponding to more "precise" words. With this analogy in mind, one can consider the quantity $p(E,t)$ as the probability of finding the system at energy E at time t.

We can write the dynamics in continuous time, by saying that if the system is initially at an energy level E, it chooses a site E' with lower energy and goes there with a probability $\Gamma \mathrm{d}t$, and so the master equation reads

$$\partial_t p\,(E,t) = -\Gamma p(E) + \Gamma \int_E^{\infty} \mathrm{d}E'\, \frac{p(E')}{\mathscr{P}_<(E')}. \tag{3.10}$$

Hence, Γ is the inverse of the time required to jump from one state to the other (equivalently, the time between the uttering or writing of two words). Note that this master equation can be deduced from a transition matrix of the form

$$W(E \to E') = \frac{\mathbf{1}_{E'<E}}{\mathscr{P}_<(E)}, \tag{3.11}$$

stating that the new site is picked with uniform probability among those that verify $E' < E$.

1 Note that the initial model is entirely discrete, and one has only N sets $\Omega_0 \supset \ldots \supset \Omega_N$. As soon as one reaches the smallest subset Ω_N then the process starts again at Ω_0. This version allows a continuous treatment of the model, with the same conclusions.

Intuitively, it is immediate that Eq. (3.10) has no stationary state: the system goes down into lower and lower energy levels. In fact, imagining that one follows the evolution of the system by measuring the corresponding energy $E(t)$ at time t, one can introduce a variable $u(t) = \mathscr{P}_<(E(t))$, and notice that at each iteration it is multiplied by a random variable with a uniform distribution in $[0;1]$.

Taking a continuous time limit, the random variable $u(t+\mathrm{d}t)$ is equal to $a(t)u(t)$ with probability $\Gamma \mathrm{d}t$ and with $a(t)$ is random variable with a uniform distribution in $[0;1]$, and to $u(t)$ with probability $1-\Gamma \mathrm{d}t$. Hence, the quantity $\ell(t) = -\ln(u(t))$ follows an additive random walk, and the Fokker-Planck equation describing the evolution of its probability density $p(\ell,t)$ can be written as

$$\frac{\partial p(\ell,t)}{\partial t} = -\Gamma \frac{\partial p(\ell,t)}{\partial \ell} + \frac{\Gamma}{2}\frac{\partial^2 p(\ell,t)}{\partial \ell^2}, \tag{3.12}$$

which is a heat equation. (See Appendix B for a quick introduction to Fokker-Planck equations).

This is fairly standard, and can be solved for $p(u,t)$, the density of the variable u, as

$$p(u,t) = \frac{1}{u}\int \mathrm{d}\xi \, \frac{M_0(\xi)}{\sqrt{2\pi\Gamma t}} \exp\left(-\frac{(\ln(u)+\xi+\Gamma t)^2}{2\Gamma t}\right), \tag{3.13}$$

with M_0 the initial condition such that $u\, p(u,0) = M_0(-\ln(u))$.

Naturally one sees something that could have been easily deduced: left on its own, the process finds lower and lower energy levels, without ever settling down. Eq. (3.13) shows, for example, that at long times the probability density is non-negligible for $\mathscr{P}_<(E(t)) = u(t) \approx e^{-\Gamma t}$ as the sample space grows progressively smaller and smaller.

As a solution to this, one can imagine a "regularized" dynamics, where a sentence can finish with probability φ per unit time and lead to a new sentence starting at an arbitrary point in the sample space . This can be done by modifying Eq. (3.10), and introducing the variable $u = \mathscr{P}_<(E)$, with density $p(u,t)$:

$$\partial_t p\,(u,t) = -\Gamma p(u,t) + \Gamma \int_u^1 \frac{\mathrm{d}v}{v}\, p(v,t) - \varphi p(u,t) + \varphi. \tag{3.14}$$

The stationary state $p(u)$ can be found by setting the right-hand side to 0, as

$$-\Gamma p'(u) - \Gamma \frac{p(u)}{u} - \varphi p'(u) = 0, \tag{3.15}$$

which after introducing $\phi := \varphi/\Gamma$ and imposing $\int_0^1 \mathrm{d}u\, p(u) = 1$ reads

$$p(u) = \frac{\phi}{1+\phi} u^{-\frac{1}{1+\phi}}, \tag{3.16}$$

where the constant $\phi/(1+\phi)$ is chosen as to have $\int n = 1$.

The stationary density $n(E)$ also reads

$$p(E) = \frac{\phi}{1+\phi}\mathscr{P}_<(E)^{-\frac{1}{1+\phi}}, \tag{3.17}$$

which, in a text of length N, can be interpreted as the ratio $N(u,t)/N$ where $N(u,t)$ is the number of instances of a word corresponding to an "energy" such that $\mathscr{P}_<(E) = u$, or equivalently corresponding to a sample space of size u.

The probability $P(n)$ that any given word has been used n times is therefore given by

$$P(n) = \int \mathrm{d}E\, \mathscr{P}(E)\delta\left(n - \frac{\phi}{1+\phi}\mathscr{P}_<(E)^{-\frac{1}{1+\phi}}\right) = \phi\left(\frac{\phi}{1+\phi}\right)^{1+\phi} n^{-2-\phi}, \tag{3.18}$$

with $n > \phi/(1+\phi)$ and with $\delta(\cdot)$ the Dirac delta distribution. One finds then a Zipf law with exponent $\mu = 1+\phi$, and the same result as obtained by Corominas-Murtra et al. in [94] in a discrete setting. Interestingly, the distribution is *completely* independent of the distribution $\mathscr{P}(E)$, showing that only the "sample space reducing" character of the process is relevant to find the limiting Zipf law. I show the result of a simulation of this process with $\phi = 0.1$ in Figure 3.2, left panel.

3.1.3.1 *Some comments on SSR processes*

Some points may now be raised for this simple model. The first point is that the model seems overly sensitive to the ordering of the words. Take for example the sentence

"The wolf howled",

where the ending verb "howled" seems extremely conditioned by the subject "the wolf". However, it could be argued that the SSR model would fail to explain the Zipf law in languages with a different syntax, as in e. g. German, where the sentence

"Der Pianist hat gestern gespielt",

can be translated to "the pianist played yesterday", but has a past participle "gespielt" (played) at the end of the sentence, even though it could be argued that the corresponding verb "spielen" (to play) was strongly constrained by the use of the subject "der Pianist". Under this model, the sentence formation would be different. Even though superficially the sentence has the same beginning as

"Der Pianist hat ein Klavier",

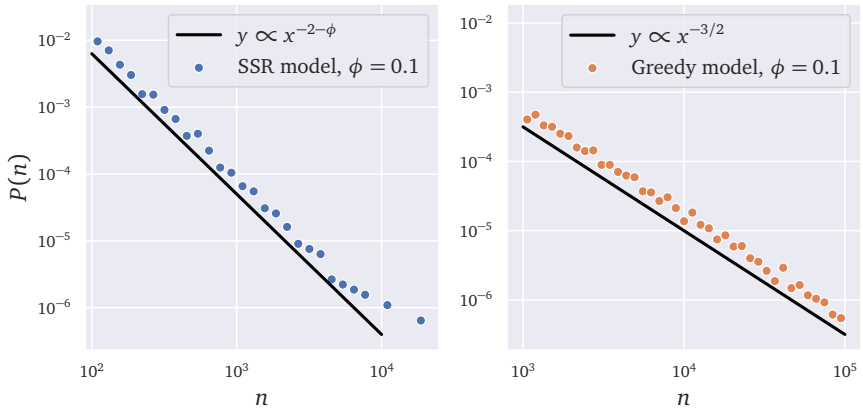


Figure 3.2: An illustration of the "sizes" distributions $P(n)$ both for the SSR model described in Sec. 3.1.3 and the Greedy model described in Sec. 3.2. Note that a Zipf law is retrieved in the SSR case, with $\mu = 1 + \phi$ and $\phi \ll 1$. On the other hand, the exponent for the Greedy model is always $\mu = 3/2$ independently of the value of ϕ.

meaning "the pianist has a piano", these two sentences were formed differently, in that "hat [...] gespielt" (played) constrained the adverb "gestern" (yesterday), while in the second sentence the word "hat" (has) constrained the choice of a direct object "a piano".

Hanel and Thurner [138] proposed a way to amend this, through so-called grammar-ordered SSR, where a sentence is first formed globally using a sample-space reducing process, with subsequent local re-arrangements of the words to preserve a correct syntax. Through this reshuffling the resulting frequency distribution is unchanged, while keeping grammatical and syntactic coherence.

The second point I should like to stress is that the model, which can in fact be modified to produce a large set of distributions, was initially conceived as an alternative to existing generative models for power-laws which rested mainly on a hypothesis of multiplicative growth. Although it is not the case here *per se*, the dynamics in Eq. (3.12) show that we are indeed not very far.

In the next section, I will explore an extension to these models.

3.2 A GENERIC MODEL: GREEDY ALGORITHMS

The following model can be imagined as a modified SSR reduced model, and was proposed as an attempt to reproduce power-laws in city or firm sizes with a plausible mechanism. It's the continuation of the continuous version of the SSR model exposed in the previous section, and an original contribution by Jean-Philippe Bouchaud and myself to these models, published in [190].

It was initially inspired by mechanisms proposed by Gualdi & Mandel [136] and Axtell [26] to explain power-law distributions describing the

size of firms defined by their number of employees. The main idea is that larger firms are more "attractive" than smaller firms, and individual agents tend to switch from small firms to large firms because of that attractiveness.

In the model, agents randomly select over time better and better items, possibly using multiple criteria. For definiteness, I will talk about firms, but other interpretations are possible. I postulate that each firm i is characterized by a set of K attributes, with scores $\mathbf{x}^i \in [0,1]^K$ that all agents agree on. For instance, x_1^i is the score given to the wages paid by firm i, x_2^i the score given to its location, x_3^i the score given to its work environment, etc. Firm j is deemed "better" than firm i when the number of attributes for which j has a higher score than i exceeds M (with $1 \leq M \leq K$).

The "algorithm" followed by agents in this case is that an agent initially in a firm characterized by $\vec{x} = (x_1, \ldots, x_K)$ will first select another firm $\vec{y} = (y_1, \ldots, y_K)$ randomly, and then decide to change to that firm only if the new firm is "better" than the old one, and staying in the latter if that isn't the case.

In the case $K = 1$ the model has strong similarities with the SSR model, with the fact of picking better and better choices mimicking the idea of a progressively reduced set of choices. The key difference is however that in the SSR model a better choice is picked *immediately*, whereas in this modified version the agent must first find a better firm before choosing it, a process that gets slower and slower in time.

Introducing a variable $u = 1 - x$ to keep the similarity with the previous model, one may write a master equation for the process, with $n(u,t)$ the density of the variable u at time t, proportional to the number of agents employed in firms with "scores" between u and $u + \mathrm{d}u$, as

$$\frac{\partial n(u,t)}{\partial t} = -\Gamma u n(u,t) + \Gamma \int_u^1 \mathrm{d}v \; n(v,t). \tag{3.19}$$

This is in fact the master equation corresponding to the 0 temperature dynamics of the so-called Barrat-Mézard model [37], where a particle moves in a landscape of randomly distributed energy levels. The mapping is as before, with $u = \mathscr{P}_<(E)$ the cdf of the energy levels.

The full solution for Eq. (3.19) may be found using Laplace transforms and the method of characteristics, as

$$n(u,t) = e^{-\Gamma u t} n(u,0) + \Gamma t e^{-u\Gamma t} \int_u^1 \mathrm{d}v \; n(v,0), \tag{3.20}$$

and one sees again that there is no stationary state.

This solution becomes universal at large times in the regime where $u = z/\Gamma t$, z finite. One finds:

$$n(u,t) = \Gamma t \mathscr{F}(u\Gamma t); \qquad \mathscr{F}(z) = e^{-z}. \tag{3.21}$$

It's also possible to compute the distribution $p_t(\tau)$ of trapping times $\tau := \frac{1}{\Gamma u}$, i. e. the time required to find a better firm after the system has evolved for a time t, given by:

$$p_t(\tau) = \frac{t}{\tau^2} e^{-\frac{t}{\tau}} \tag{3.22}$$

in full agreement with the findings of [37] in the appropriate limit $t \to \infty$, $\tau \to \infty$ with $t/\tau = O(1)$. This is a first manifestation of a phenomenon called *aging*, which shall be discussed later in Chapter 8.

As in the previous section, one may "regularize" this model, by saying that a firm may go bankrupt at a rate φ, in which case its employees are randomly distributed among all other firms. The new master equation now reads

$$\frac{\partial n(u,t)}{\partial t} = -\Gamma u n(u,t) + \Gamma \int_u^1 \mathrm{d}v \; n(v,t) - \varphi n(u,t) + \varphi. \tag{3.23}$$

Its stationary state $n(u)$ can be found by setting the right hand side of Eq. (3.23) to 0 and differentiating, leading to a normalized solution

$$n(u) = \frac{\phi(\phi+1)}{(u+\phi)^2}, \tag{3.24}$$

with $\phi = \frac{\varphi}{\Gamma}$.

The full dynamical solution can also be obtained,

$$n(u,t) = n(u) + e^{-(u+\phi)\Gamma t} m(u,0) + \Gamma t e^{-(u+\phi)\Gamma t} \int_u^1 \mathrm{d}v \, m(v,0), \tag{3.25}$$

where $m(u,0) = n(u,0) - n(u)$. Hence, the system converges to its stationary state after a time $\sim (u\Gamma)^{-1}$ for $u \gg \phi^{-1}$ and after a time $\sim \varphi^{-1}$ for small u's. Hence, ϕ^{-1} acts as both a cut-off limiting the values attainable by s, and as the relaxation time-scale of the system.

With the stationary solution at hand for $n(u)$, one can find the firm "size" distribution, using the same formula as before

$$P(n) = \int_0^1 \mathrm{d}u \, \delta\left(n - \frac{\phi(\phi+1)}{(u+\phi)^2}\right) = \frac{\sqrt{\phi(1+\phi)}}{2} n^{-3/2}, \tag{3.26}$$

holding in a range $n \in [\phi/(1+\phi), 1+1/\phi]$, outside of which $P(n)$ is equal to zero. I've found then an exponent $\mu = 1/2$ and therefore a distribution that is *fatter* than for the SSR process. Additionally, the reader has now come across the first instance of a mechanism for power-laws where the exponent does *not* depend on any of the parameters of the model, a feature that will emerge later in other so called "critical" models. A depiction of this distribution from a simulation with $\phi = 0.1$ is shown in Figure 3.2, right panel.

This in fact suffices to describe the main features of the model. The reader may find below the discussion for the multi-criterion case $K > 1$.

3.2.1 *A few generalisations*

I shall first illustrate the difference between the standard SSR from Sec. 3.1.3 and the previous "greedy" model. It can be be seen that the difference between the standard SSR models and the "greedy algorithm" above is chiefly that for the first case one always picks a site with a lower energy and moves to it with probability 1, so that the system is constantly switching. On the other hand, on the greedy side it is necessary to first *find* a site with lower energy, a process that gets slower and slower as the time required for that is of the order of u^{-1}.

In fact the two dynamics in Eq. (3.10) and (3.19), as their solutions, are equivalent up to the change of variables:

$$n(u) \propto p(u)/u, \tag{3.27}$$

showing the influence of this "waiting time" u^{-1}.

In fact the greedy model can be itself mapped into a problem where a certain choice u has been made, and then a better choice $v < u$ is made but *only* after waiting a time $1/u$. If one instead waited a time $1/u^{\beta}$, with $\beta > -1$, the stationary distribution would read

$$n_{\beta}(u) \propto u^{-\beta-1}, \tag{3.28}$$

and therefore lead to a size distribution

$$P(n) \propto n^{-\frac{\beta+2}{\beta+1}} \tag{3.29}$$

allowing then for a host of exponents, indifferently of the distribution $\mathscr{P}(E)$.

In the multi-criterion case, there are instead K coordinates $(u_1, \ldots, u_K) := \vec{u}$, and a new site $\vec{v}$ may only be picked if there are M indices $i_1, \ldots, i_M$ such that $v_{i_l} < u_{i_l}$. I denote a matrix $W_{\vec{u}\to\vec{v}}$ that is equal to 1 when this condition is satisfied and 0 when it isn't.

In this case, the master equation can be written

$$\frac{\partial n(\vec{u},t)}{\partial t} = -\Gamma n(\vec{u},t)\omega_{K,M}^{\beta}(\vec{u}) + \Gamma \int_u^1 \mathrm{d}\vec{v}\, n(\vec{v},t)\omega_{K,M}^{\beta-1}(\vec{u}) - \varphi n(\vec{u},t) + \varphi, \tag{3.30}$$

with $\beta = 1$ in the original "greedy" dynamics described above, and where

$$\omega_{K,M}(\vec{u}) = \int_{[0,1]^K} \mathrm{d}\vec{v}\, W_{\vec{u}\to\vec{v}}. \tag{3.31}$$

This last quantity corresponds to the volume of the space of coordinates $\vec{v}$ that are "better" than $\vec{u}$. The master equation Eq. (3.30) corresponds then to a generalized dynamics that is equivalent to being at a site $\vec{u}$, waiting

a time $\omega_{K,M}(\vec{u})^{-1}$ and then selecting a new coordinate $\vec{v}$ that is "better" than $\vec{u}$ in the sense defined above.

In [190] it was shown that the volume defined in Eq. (3.31) can be computed using spin variables in the limit $M, K \to \infty$ with $\alpha := M/K$ fixed. The model has a transition at $\alpha = 1/2$: above this value, $\omega_{K,M}(\vec{u})$ diminishes as the system evolves, and in the case $\beta = 0$, corresponding to the SSR model when $K = 1$, the model leads to a Zipf law as before, with the distribution defined in Eq. (3.29) becoming instead

$$P(n) \propto \begin{cases} n^{-2} F_K(n) & \text{for} \quad n < n^*(K) \\ 0 & \text{for} \quad n > n^*(K), \end{cases} \tag{3.32}$$

where F_K is a piece-wise polynomial in $\ln(n)$ and $n^*(K)$ a cut-off that grows exponentially with K. Interestingly, when $\frac{1}{2} < \alpha < 1$ there is no need to introduce a "regularization" term φ as before. Note that this isn't a perfect power-law because of the term $F_K(n)$, but it would be very difficult to distinguish it from a true one in practice.

3.3 STOCHASTIC MULTIPLICATIVE GROWTH MODELS

The section below discusses a very generic mechanism for the origin of power-laws, linked with what I have shown previously. These models consist chiefly of multiplicative growth, where a quantity grows by a random factor proportional to the quantity itself. I will first introduce the Yule and Barabási-Albert preferential attachment model, as they are very closely linked to the Simon model described above. I will then show how "bare" multiplicative growth leads not to a power-law, but to a log-normal distribution, but that this is not something that is heuristically very far off from a power-law. This is close to the idea, first enunciated to my knowledge by Levy and Solomon, that power-laws are the generic result of multiplicative processes. I will then show how generically adding "frictions" to multiplicative growth leads to power-laws, presenting also analytic arguments to explain this fact. This section relies heavily on the reviews by Simkin and Roychowdhury [240] and by Kumamoto and Kamihigashi [162].

3.3.1 *The Yule model*

In the 1920s, Willis, seeking to enlighten the mechanisms that drive the evolution of species, compiled data on the statistical distribution of the frequency of biological genera [275]. He and Yule then found that this quantity appeared to be power-law distributed [276], motivating Yule to propose the following model to explain this [280].

In this model there are two main entities: species and genera. The analogy with language is that each species is an instance of a given genus, as an item in a text is an instance of a general word. In other words,

the model is strictly analogous to Simon's model for the word frequency distribution in natural language.

In Yule's model, a new species is added to a genera with a probability per unit time proportional to s, and a mutation leading to a new genus appears with a rate g. Quite straightforwardly, one sees that the number of different genera scales on average as $\exp(gt)$, and so the frequency of genera with age t is given by

$$p(t) = ge^{-gt}. \tag{3.33}$$

On the other hand, a genus of age t has on average

$$n(t) = \exp(st) \tag{3.34}$$

species. Thus if one considers the distribution $p(n)$ of n one is simply looking at the exponential of an exponentially distributed random variable, as given in Eq. (3.33). One can then heuristically write $p(n)\mathrm{d}n = p(t)\mathrm{d}t$, with $t = \ln(n)/s$, to find

$$p(n) \propto n^{-1-\mu}, \quad \mu = \frac{g}{s}, \tag{3.35}$$

which furthermore shows the general result that the exponential of an exponentially distributed random variable X with density $p(x) = \mu e^{-\mu x}$ is a power-law distribution of exponent μ. Explicitly here the "friction" mechanism is given by g, as when $g > 0$ it simply states that the probability of finding a genus of age t decreases exponentially with t. This very simple explanation therefore leads to a power-law distribution, but one may also map Simon's model as presented in Sec. 3.1.2 directly onto this model.

Indeed, if one denotes by k the number of appearances of a given word in a text, while calling N the total number of words, as specified in Sec. 3.1.2, one can also introduce a variable D for the number of *distinct* words and do the following analysis. As the text is being increased by one word at each time-step, during a step $\Delta N = 1$, the probability per unit time to increase the number of appearances of a word from k to $k+1$ is proportional to k/N, and given by

$$P(k \to k+1) = (1-\alpha)\, k \frac{\Delta N}{N} \approx (1-\alpha) k \Delta t \tag{3.36}$$

where $t = \ln(N)$. This means that on average one expects the number of appearances of a given word, the first appearance of which appeared N words ago, to be given by $N^{1-\alpha} = \exp\left((1-\alpha)t\right)$ under the mapping $t = \ln(N)$. This is equivalent to the number of species within a genus of age t, and so one maps $s = 1-\alpha$, where s is the intra-genus mutation rate in the Yule model.

Similarly, one expects the number of distinct words to increase on average by α during a time-step Δt, and so

$$P(D \to D+1) = \alpha \Delta N \approx \alpha N \Delta t, \tag{3.37}$$

and so one sees that here α plays the role of g, as the rate with which new genera, or new words, appear, and $1-\alpha$ plays the role of s as the intra-genus mutation rate. With this, one readily recovers the equivalence

$$\mu = 1 + \frac{g}{s} = \frac{1}{1-\alpha} \tag{3.38}$$

for the Simon model, which now appears also clearly a mixture of exponentials, to which I shall return later.

Interestingly, this same mechanism was used by Enrico Fermi to explain the empirical observation stating that the energy spectrum of cosmic radiation is well described by a power-law, see [110].

3.3.2 *Network growth and preferential attachment*

The previous Yule-Simon family of models can in fact also be transposed into a network-science setting, as a specific case of the growing random network described in e. g. [160].

A graph or network is a collection of vertices (also called nodes) V, labelled for example $i = 1, \ldots, N$, and edges E, represented as $i \longleftrightarrow j$ if the network is undirected or $i \rightarrow j$ if the network is directed. An example (one hopes, at least!) of an undirected network is that of friendship, with the vertices representing people and the edge $i \longleftrightarrow j$ defined in the network if i and j are friends; and an example of a directed network is the network of citations of scientific articles, where each node is a paper and edge $i \rightarrow j$ belonging to the network if i cites j.

The statistical study of networks then focuses on a few key quantities. For example, one can define the in-degree of a node i as the number of links $j \rightarrow i$ in the graph, and similarly for the out-degree, where one counts instead edges $i \rightarrow j$ (for an undirected graph the extension is natural, and one speaks only of the degree). Similarly, one can focus on "connected components", that is a sub-network such that any two nodes i and j in it can be linked by a path $i \rightarrow i_1 \rightarrow \ldots \rightarrow j$, with the caveat that one speaks of a weakly connected component if any such path can only be gone down in one way, i. e. from i to j and not vice-versa.

Many networks are "scale-free", in the sense that the quantities described above are often power-law distributed, and very generic models of "growing random networks" have been created to explain this. In a nutshell, these models imagine that a network grows organically, adding for example one node at a time. This idea is represented in Figure 3.3, where a single node has just been added to the network and one has many possibilities to add an edge from it to any other existing vertex. Such dynamics can be used, for example, to explain power-law distributed sizes of clusters. Saying that the new node attaches to an existing cluster with probability $1-\alpha$ and creates a new one with probability α induces Simon-like dynamics for the clusters and therefore a power-law distribution in their size, as we've seen previously in Sec. 3.1.2 and Sec. 3.3.1. A very famous model along this lines is the preferential attachment model by Barabási and Albert [5].

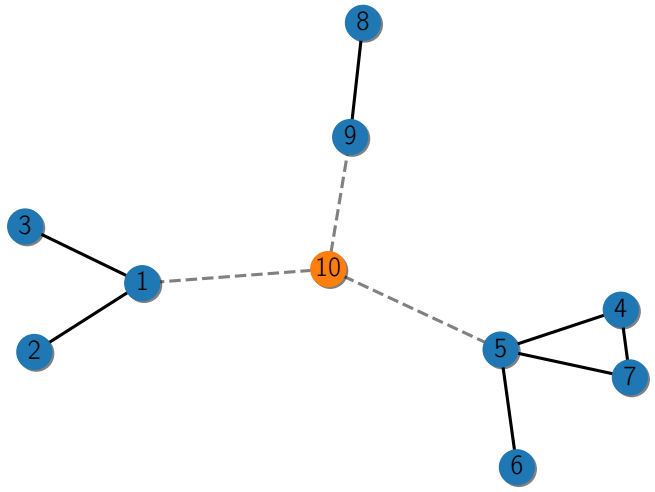


Figure 3.3: An illustration of a random growing undirected network: there are three "clusters" or connected components, each containing the nodes $(1,2,3)$, $(4,5,6,7)$ and $(8,9)$. A new node labelled 10 has just been added to the network, and may attach itself to other nodes by developing an edge to any existing node, as exemplified by the dashed edges (although we've only represented one dashed edge per connected component).
The precise stochastic rules with which this happens are determined by the specifics of the model, but in general one may be interested in the distribution of the sizes of the connected components. Here, for example, the sizes are 3, 4 and 2, and one of these will grow by 1 once the new node is added. One may also be interested in the distribution of degrees, defined as the number of edges attached to a node. Here, for example, the node labelled 5 has degree 3.

3.3.2.1 *The Barabási-Albert model*

The motivation behind this model is to explain power-law distributed degree distributions. As an analogy, it can be thought of as an attempt to explaining the distribution of the number of citations a given scientific article may have. It is similar in concept to the Simon model, but the main difference is that we are interested in the *degree* distribution in the network, and not in the size of clusters as before. The dynamics is also different: instead of having the new vertex latch onto any existing node with a probability $1-\alpha$, it attaches instead to an existing node with a probability that is proportional to its degree.

If one starts with 1 node and no edges, then each time step allows one to add one edge. After t steps there are therefore t edges, and so the sum of the degrees of all nodes is $2t$. As for Eq. (3.4) one may then write an

equation for $N(k,t)$, the number of nodes with degree k at time t, that reads

$$N(k,t+1)-N(k,t)=\frac{k-1}{2t}N(k-1,t)-\frac{k}{2t}N(k,t), \tag{3.39}$$

as $k/2t$ is the probability of attaching the new node to a node of degree k.

As before, one may take the continuous limit when t is large for the quantity $P(k,t)=N(k,t)/t$, which is the proportion of nodes with degree k. The equation obtained is

$$\frac{\partial\left[tP(k,t)\right]}{\partial t}=\frac{1}{2}\left[(k-1)P(k-1,t)-kP(k,t)\right]+\delta_{k,1}, \tag{3.40}$$

and one can try to find the stationary state by setting $\partial_t P(k,t)=0$. This can be done exactly, but I resort again to a continuum approximation, as for the stationary distribution of the Simon model in Eq. (3.6), obtaining

$$\frac{1}{2}\frac{\mathrm{d}\left[kP(k)\right]}{\mathrm{d}k}=-P(k), \tag{3.41}$$

which solves as $P(k)\propto k^{-3}$, and corresponds therefore to $\mu=3$. It is also equivalent when $k\gg 1$ to the exact solution, given by

$$P(k)=\frac{4}{k(k+1)(k+2)}. \tag{3.42}$$

Note that this model corresponds to a *linear* attachment probability, as the probability to attach a new vertex to a vertex with degree k is proportional to k. Other models, where this probability is $\propto k^{\gamma}$, may be studied and lead to other exponents, see Ref. [160]. I should also like to add that the study of the evolution of networks is a huge and extremely interesting topic, and that the degree distribution is largely insufficient to capture all the necessary details in a network, see Ref. [102] for a good review on generative models and other aspects of networks.

Finally, I should like to note that this model is equivalent to the Yule model, and therefore also to the Simon model, with $g=s$, as noted by Simkin and Roychowdhury [240]. The equivalent of the genus is a node, and the number of species within each node is the equivalent of the node's degree, and it is clear that in this model the rates of appearances of an edge and a node are the same.

3.3.3 *Frictionless multiplicative growth and Derrida's Random Energy Model*

In some of the previous models the quantity of interest, be it the remaining available "phase-space" $\mathscr{P}_<(E)$ from Sec. (3.1.3) or directly the quantities denoted by k (word frequencies, sizes of genera, etc...) in the previous sections, evolved multiplicatively. Switching to a notation x_t for any of these quantities at time t, the dynamics were often of the form

$$x_{t+1}=g_t x_t, \tag{3.43}$$

where g_t encodes the random fluctuations governing the evolution of x_t at time t. If one introduces instead the variable $\ell_t = \ln(x_t)$, and calls $g_t = e^{\eta_t}$, then these multiplicative dynamics become additive dynamics, and read

$$\ell_{t+1} = \ell_t + \eta_t. \tag{3.44}$$

Considering then that the η variables are i.i.d., and writing $m = \mathbb{E}[\eta]$ and $\sigma^2 = \mathbb{V}[\eta]$, leads to Fokker-Planck equations for the probability densities of the two variables, $p(x,t)$ and $p(\ell,t)$, that read

$$\frac{\partial [p(x,t)]}{\partial t} = -\frac{\partial}{\partial x}\left[\left(m + \frac{\sigma^2}{2}\right) x p(x,t)\right] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}\left[x^2 p(x,t)\right] \tag{3.45}$$

for x, and

$$\frac{\partial [p(\ell,t)]}{\partial t} = -m\frac{\partial [p(\ell,t)]}{\partial \ell} + \frac{\sigma^2}{2}\frac{\partial^2 [p(\ell,t)]}{\partial \ell^2} \tag{3.46}$$

for ℓ. The second equation, a heat equation, is much more amenable, and can be solved directly for e. g. an initial condition $\ell_0 = 0$, equivalent to $p(\ell,0) = \delta(\ell)$. The solution reads

$$p(\ell,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(\ell - mt)^2}{2\sigma^2 t}\right), \tag{3.47}$$

a gaussian distribution. This could of course have been guessed from the form of Eq. (3.44), where one should expect to have $\ell_t = \sum_{t'=0}^{t-1} \eta_t$, and so to converge to a gaussian of mean mt and variance $\sigma^2 t$ *via* the central limit theorem.

This solution can then be used to find the solution of Eq. (3.45) by changing variables. The solution to that equation reads then

$$p(x,t) = \frac{1}{x\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(\ln(x) - mt)^2}{2\sigma^2 t}\right), \tag{3.48}$$

which is now a *log-normal* distribution. As a result, one may represent the variable x_t as

$$x_t = \exp\left(mt + \sqrt{t}\sigma\xi\right), \tag{3.49}$$

where ξ is a gaussian random variable with unit variance.

The reader may now wonder what has happened, as I have spent the previous sections showing that multiplicative growth tends to lead to power-law distributions, and not to a log-normal. My first reply to this should be that in this case there is no *stationary* state, contrary to what has happened previously. The distribution never settles, and statistical equilibrium is never reached. For example, despite having well defined mean and variance, these quantities diverge, as

$$\mathbb{E}[x(t)] = e^{mt + \frac{\sigma^2}{2}t}, \quad \mathbb{V}[x(t)] = [e^{\sigma^2 t} - 1]e^{2mt + \frac{\sigma^2}{2}t}. \tag{3.50}$$

Interestingly, the mode or most likely value of the distribution behaves as $\exp\big((m-\sigma^2)t\big)$, and one can end up in a situation where the mean diverges to $+\infty$, while the most likely value goes to 0!

The second point I wish to raise is that this distribution is in fact not too far from a power-law distribution. Indeed, expanding the terms inside the exponential in Eq. (3.48) leads immediately to

$$p(x,t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{m^2}{2\sigma^2}t} x^{-1-\mu(x,t)}, \quad \mu(x,t) = -\frac{m}{\sigma^2} + \frac{\ln(x)}{2\sigma^2 t}, \tag{3.51}$$

and so when $2\sigma^2 t \gg \ln(x)$ the distribution resembles a power-law with an exponential cut-off.

3.3.3.1 *The Random Energy Model*

A good deal of things can be studied with this simplistic multiplicative growth model. Going back to the discussion in Chapter 1, one may think of the quantity x_t as the wealth of an individual, and considering then the evolution of the wealth of N individuals by considering N realizations of the process in Eq. (3.43). The wealth of the i-th individual is therefore denoted $x_{i,t}$.

I now study how this wealth is distributed among the N individuals by considering the share $w_{i,t}$ of the total wealth owned by an individual. Owing to Eq. (3.49), this quantity reads

$$w_{i,t} = \frac{\exp\left(\sigma\sqrt{t}\xi_i\right)}{\sum_j \exp\left(\sigma\sqrt{t}\xi_j\right)}, \tag{3.52}$$

and its study is therefore equivalent to that of Derrida's Random Energy Model [99, 100].

The model was initially introduced as a way to understand the qualitative behaviour of the Sherrington-Kirkpatrick (S.K.) model for spin-glasses [238], as a formal limit of the so-called p-spin model. Within that model, the probability attributed to a configuration $\vec{s} = (s_1, \ldots, s_N)$ of spin variables, with $s_i = \pm 1$, is given by

$$p(\vec{s}) \propto e^{-\beta \mathscr{H}(\vec{s})}, \quad \mathscr{H}(\vec{s}) = \sum_{i_k} J_{i_1,\ldots,i_p} s_{i_1} \ldots s_{i_p}, \tag{3.53}$$

where β is an inverse temperature. In the limit $p \to \infty$, Derrida showed that the study of this model was equivalent to studying the thermodynamic properties of a system with N energy levels E_i drawn from a gaussian distribution, so that the probability of finding the system in one of the energy levels E_i is given by

$$p(E_i) = \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}}, \tag{3.54}$$

which is exactly of the form given in Eq. (3.52) if one sets $\beta = \sigma\sqrt{t}$ (since $-E_i$ and ξ_i are statistically equivalent).

Figure 3.4: A numerical experiment for the pure multiplicative growth model. Here, the wealth $x_{i,t}$ of an individual grows as $x_{i,t+1} = e^{\xi_{i,t}} x_{i,t}$, where ξ is a centred gaussian variable of unit variance. There are ten individuals, and the top panel shows the share each individual possesses of the total wealth, $w_{i,t} = x_{i,t} / \sum x_{j,t}$. The bottom panel shows the evolution of the herfindahl index averaged over a hundred such experiments. As discussed in Sec. 2.3.1, the herfindahl should be equal to $1/10$ in the case of perfect equality among the ten individuals but equal to 1 when a single individual captures all the wealth. As shown by the Random Energy Model, this very simplistic model invariably leads to condensation.

Given the dichotomy I stated above between the behaviour of the mode and the mean of a log-normal distribution, the reader may intuitively expect this measure to condense as $t \to \infty$. This is indeed what happens, and to understand it I reformulate the argument given in [65] by focusing again on extreme values.

One wishes therefore to study the large values of $x_{i,t}$, and these should naturally correspond to the large values of ξ_i. Indeed, as sketched in Sec. 2.2.1, when $N \gg 1$, the maximum $\xi_{(0)}$ of these variables can be written as

$$\xi_{(0)} \approx \sqrt{2\ln(N)} + \frac{1}{\sqrt{2\ln(N)}}\, \upsilon, \tag{3.55}$$

where υ is a Gumbel random variable,[2] and therefore the tails of its distribution are given by $p(\upsilon) \underset{\upsilon \gg 1}{\approx} e^{-\upsilon}$.

For the largest share $w_{(0),t}$ one may write

$$w_{(0),t} \propto \exp\left(\sigma \sqrt{\frac{t}{2\ln(N)}}\, \upsilon\right), \tag{3.56}$$

an exponential of a random variable that is asymptotically exponential, as in Sec. 3.3.1. Hence, the tails of the distribution of w behave as

$$p(w) \propto w^{-1-\mu}, \quad \mu = \sqrt{\frac{2\ln(N)}{\sigma^2 t}}, \tag{3.57}$$

and, using the same arguments as in Sec. 2.3.1, the reader should expect a condensation transition in the limit $t, N \to \infty$ with fixed $\mu = \frac{2\ln(N)}{\sigma^2 t} = 1$, corresponding to the transition temperature found by Derrida in [99, 100].

As usual this can be formulated more precisely, as done in Ref. [42], by considering the sum in the denominator of Eq. (3.52), equivalent to the partition function of the Random Energy Model,

$$Z_N(t) = \sum_{j=1}^{N} \exp\left(\sigma \sqrt{t} \xi_j\right). \tag{3.58}$$

Indeed, in the limit described above, $t, N \to \infty$ with $\mu = \frac{2\ln(N)}{\sigma^2 t}$ fixed, it can be proved that

$$Z_N(t) \propto \begin{cases} \chi_G & \text{if} \quad \mu > 2, \\ \chi_L & \text{if} \quad \mu \leq 2 \end{cases}, \tag{3.59}$$

where χ_G is a gaussian random variable, and χ_L a Lévy-stable variable with tail parameter μ.

In definiteness, this model of frictionless multiplicative growth for the wealth of N individuals can be thought of as being described by a power-law distribution, albeit one with a tail that gets fatter and fatter as time increases. A stunning consequence of this, as represented in Fig. 3.4, is that this "wealth condensation" happens mechanically, and a single individual captures all the wealth out of sheer luck. Another interesting aspect of the study of this model is that it hints at something I shall study in Chapters 7 and 8: ergodicity. If one runs the model for N individuals up to a certain time T, then taking fixed N and increasing T leads to a more and more unequal result, while doing the opposite by fixing T and increasing N progressively will do just the opposite. The two "longitudinal" and "cross-sectional" ways of averaging do not commute in this case!

2 Note that only the dominant deterministic term $\sqrt{2\ln(N)}$ has been written explicitly. The sub-leading is formally larger than $1/\sqrt{2\ln(N)}$, and equals $\gamma\sqrt{\frac{2}{\ln(N)}}$ with γ the Euler-Mascheroni constant. Note that $\gamma\sqrt{2} \approx 0.8$.

3.3.4 *Multiplicative growth with death*

In what has been previously said, I expect that the reader has acquired the intuition that pure multiplicative growth is not too far from leading to a power-law distribution. It would seem natural to say that adding some sort of "friction" could help in this: taking Eq. (3.51), the reader could be tempted to replace the term in $\ln(x)$ by a constant, because of the very slow variation of the logarithm, and take a limit $t \to \infty$ to get a constant exponent μ. Of course, this is just a way to prevent the tail of the log-normal distribution from getting fatter and fatter, or equivalently a way of preventing the underlying quantity x_t from getting too small or too large (for definiteness, you may think of it again as wealth, but other interpretations are of course possible).

A very naive, but realistic, way of achieving this is to introduce births and deaths into the model. Introduce then the possibility that an agent may die at each time step with a probability proportional to a number λ, and is immediately replaced by a new agent with a certain wealth x^*, which I take for simplicity to be equal to 1. The Fokker-Planck equation is modified from that in Eq. (3.45), to become

$$\begin{aligned}\frac{\partial\,[p(x,t)]}{\partial t} =& -\frac{\partial}{\partial x}\left[\left(m+\frac{\sigma^2}{2}\right)xp(x,t)\right]+\frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}\left[x^2p(x,t)\right]\\ &-\lambda p(x,t)+\lambda\delta(x-x^*).\end{aligned} \tag{3.60}$$

I could attempt to solve this equation directly, but there is a more intelligent way to find the stationary distribution. Consider indeed the "age" t of a single agent, and state that the distribution $p(x|t)$ of his wealth given that age is given by the log-normal distribution in Eq. (3.48). One need only remark then that the ages of all the agents are determined by a Poissonian birth process, so that their age distribution is $p(t) = \lambda e^{-\lambda t}$. The stationary distribution can then be computed as

$$\begin{aligned}p(x) =& \int \mathrm{d}t\; p(x|t)p(t)\\ =&\frac{1}{x}\frac{1}{\sqrt{2\pi\sigma^2}}\int_0^\infty \mathrm{d}t\,\frac{\lambda}{\sqrt{t}}\exp\left(-\frac{(\ln(x)-mt)^2}{2\sigma^2 t}-\lambda t\right)\propto x^{-1-\mu},\end{aligned} \tag{3.61}$$

an exact power-law, with an exponent μ found to be

$$\mu = -\frac{m}{\sigma^2}+\frac{\sqrt{m^2+2\lambda\sigma^2}}{\sigma^2}. \tag{3.62}$$

3.3.5 *Multiplicative growth with a reflective barrier*

Another hint at the possibility of obtaining a power-law distribution for the quantity x is to get an exponential distribution for $\ell = \ln(x)$. This is

essentially the mechanism proposed in [247] with the "friction" mechanism in this case being a reflective barrier, and I sketch the intuition and its solution below.

Taking Eq. (3.46) one could indeed be tempted to suppose the existence of a stationary state, and take $\partial_t p(\ell, t) = 0$ to find the differential equation

$$0 = -mp'(\ell) + \frac{\sigma^2}{2} p''(\ell), \tag{3.63}$$

with an obvious solution $p(\ell) \propto \exp\left(\frac{2m}{\sigma^2}\ell\right)$. The problem is that this solution is not normalizable, as its integral for $\ell \in \mathbb{R}$ is divergent!

The problem is easily mended if one considers the case $m < 0$ and adds the supplementary condition that x or l may not be smaller than certain bound ℓ^* and x^*. In this case, the system converges to a stationary state, with distributions

$$p(\ell) \propto \exp\left(\frac{2m}{\sigma^2}\ell\right), \quad p(x) \propto x^{-1+\frac{2m}{\sigma^2}}, \tag{3.64}$$

and so one finds again a pure power-law with this mechanism. Both distributions found above are normalizable when $m < 0$ because of the introduction of lower cut-offs x^* and ℓ^* in the integral over x or ℓ, i. e. $\int_{x^*}^{\infty} \mathrm{d}x\ p(x) < \infty$.

In addition, one gains the advantage of having a clear physical picture. The variable ℓ_t can be thought of as the position of a particle that diffuses with a negative drift, but that may not go to the left of a reflecting wall placed at ℓ^*.

This somewhat artificial mechanism was also introduced, using a different description, by Champernowne [82]. He considered income brackets $I_k = [a^k; a^{k+1}]$ with integer $k \geq 1$, and certain probability rates $w_{k\to k+1}$ and $w_{k\to k+1}$ with which individuals can progress or regress from one bracket to the other, with the addition that no individual may earn less than a. This is effectively a multiplicative model, since brackets are defined with powers of a. Champernowne then found a power-law or Pareto distribution when the corresponding drift is negative, in a way similar to that defined in this section. Crucially, he thought of the death of an individual within his model as simply having him replaced by an heir in the same income bracket. As shown previously in Sec. 3.3.4, had he dealt with such events in a different way the somewhat artificial introduction of a negative drift would not have been necessary to produce a power-law income distribution.

3.3.6 *Multiplicative growth and redistribution: the Bouchaud-Mézard model*

The previous mechanism seems indeed rather contrived in that it emerges only when the drift of the logarithm of x is negative. This is not obvious, for example, when one thinks of it as a way of explaining wealth inequality given the semi-exponential growth of the world economy in the last century.

There exists however a scenario where a very similar picture emerges naturally.

The Bouchaud-Mézard [64] model imagines N individuals, with the wealth of individual i at time t represented by $W_i(t)$. It introduces redistribution of the individuals' wealth to model commercial exchange and taxation, and so the equation describing the evolution of W_i is given by

$$\frac{\mathrm{d}W_i}{\mathrm{d}t} = mW_i + \sigma\eta_i W_i + \sum_j \left(J_{ij}W_j - J_{ji}W_i\right). \tag{3.65}$$

Here, m is the average trend and can be imagined as being the growth rate of the economy, while η_i is a random variable with unit variance, modelling the different random shocks to i's wealth with a volatility factor σ setting their intensity. The last term distinguishes the model from that of pure multiplicative growth, as the term J_{ij} controls the distributive mechanism that sets the fraction of the wealth of j that flows to i. If one thinks of W_i as the population of a city, then J_{ij} can also be thought of as migration flows between cities. A natural simplification, at least from the standpoint of mean-field models, is to consider the case $J_{ij} = J/N$, so that there is redistribution at the same rate between all the individuals. Note however that more complicated topologies can be explored both theoretically and numerically, with the same qualitative behaviour for the wealth distribution [144, 249]. Now the equation reads

$$\frac{\mathrm{d}W_i}{\mathrm{d}t} = mW_i + \sigma\eta_i W_i + J(\overline{W} - W_i), \tag{3.66}$$

where $\overline{W} = \sum W_i/N$ is the empirical average of the wealth.

As such, the model can also be interpreted as having random growth with a tax at a flat rate given by J, but with a redistribution policy where all individuals receive a share of the total wealth in the economy.

The next step to tackling this model may seem unjustified to the reader who's just read through a section on the Random Energy Model. If one takes Eq. (3.66), sums over i and divides by N then one obtains the following equation,

$$\frac{\mathrm{d}\overline{W}}{\mathrm{d}t} = m\overline{W} + \frac{\sigma}{N}\sum_i \eta_i W_i. \tag{3.67}$$

Suppose then that the variables W_i are not too broadly distributed, and in particular that the condensation transition of the Random Energy Model does not take place for any time t. With this assumption, the sum on the right-hand side goes to 0 as $N \to \infty$ at all times, and this is equivalent to stating that the empirical average is equal to the mean over different realizations of the process, $\overline{W}(t) = \mathbb{E}[W_i(t)]$. I do stress that this step will only be justified self-consistently afterwards, as the W_i variables will indeed be found to have a finite mean.

One then has directly that $\overline{W}(t) = \overline{W}(0)e^{mt}$. I now introduce the rescaled variables

$$x_i(t) = \frac{W_i(t)}{\overline{W}(t)}, \tag{3.68}$$

which is simply the wealth of an individual counted in units of the average wealth. Its dynamics, and those of its logarithm $\ell = \ln(x)$, are described by

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = J(1-x_i) + \sigma x_i \eta_i(t) \tag{3.69}$$

and

$$\frac{\mathrm{d}\ell}{\mathrm{d}t} = -\left(J + \frac{\sigma^2}{2}\right) + Je^{-\ell} + \sigma\eta(t). \tag{3.70}$$

I may now comment on these equations in light of the previous Sec. 3.3.5. First, both equations show that the redistribution factor J plays the role of a soft reflective barrier: when x or l get too small there is an additive term proportional to J that pushes either of the quantities back up. Additionally, the reader can see that the rescaling by the average wealth had the effect of introducing a negative drift in the evolution of ℓ, as shown in Eq. (3.70). On account of these two features, very similar to those introduced in Sec. (3.3.5), it is not altogether unreasonable to see that this model produces a power-law distribution.

The Fokker-Planck equation associated to Eq. (3.69) reads

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[J(1-x)p(x,t)\right] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}\left[x^2 p(x,t)\right], \tag{3.71}$$

and as usual the stationary state may be found by setting $\partial_t p(x,t) = 0$. The differential equation can be solved easily, and yields an Inverse Gamma distribution,

$$p(x) \propto \frac{\exp\left(-\frac{2J}{\sigma^2 x}\right)}{x^{1+\mu}}, \quad \mu = \frac{2J}{\sigma^2} + 1, \tag{3.72}$$

which is asymptotically a power-law.[3] I note also that it is of the same form as that of Eq. (3.64), where the equivalent of m can be read in Eq. (3.70), $-m \leftrightarrow J + \frac{\sigma^2}{2}$, further showing the similarity between the two models.

In fact, the Bouchaud-Mézard dynamics can be considered, via the Langevin formalism described in Appendix B.1.3, as the dynamics of a particle with position ℓ in a potential $V(\ell) = \left(J + \frac{\sigma^2}{2}\right)\ell + Je^{-\ell}$. This is very similar to a potential that is linear for $\ell > \ell^*$, but with $V(\ell^*) = \infty$, which is the case of the multiplicative growth with barrier model from Sec. 3.3.5. These two potentials are depicted in Figure 3.5, and it is in both cases the linear potential that determines the exponential tails of $p(\ell)$ and therefore the power-law exponent of $p(x)$. Note that the Bouchaud-Mézard model has a potential barrier $\propto e^{-\ell}$, softer than the hard barrier of the other model but still exponentially large for $\ell < 0$.

3 The alert reader may notice a difference between my solution and that given in [64], but this is due to my favouring the Itô prescription over the Stratonovitch prescription for the stochastic differential equation in Eq. (3.69).

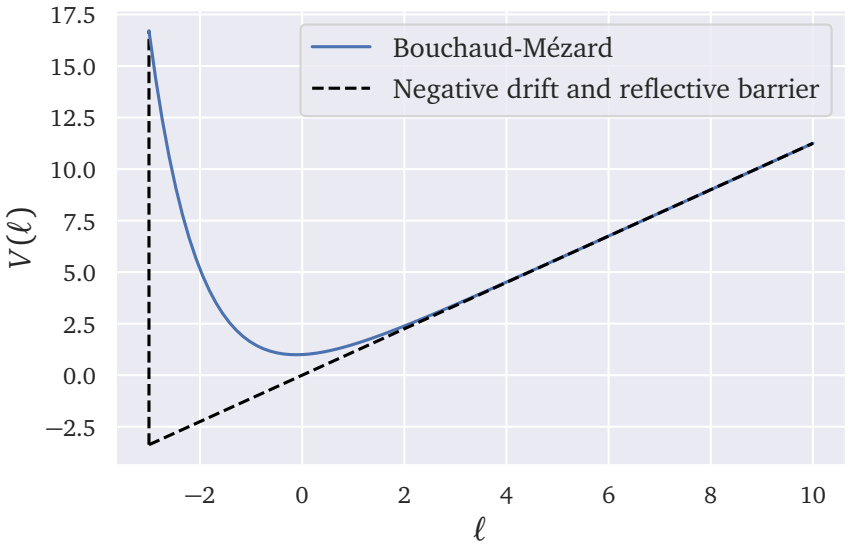


Figure 3.5: Langevin potential for $\ell = \ln(x)$ in the Bouchaud-Mézard model (solid line) and for the multiplicative growth with barrier model (dashed line). I have picked the parameters such that the linear part of the potential is the same in both models, leading to the same power-law exponent for x.

This last point is of course not very surprising. Taking indeed Eq. (3.70), its associated Fokker-Planck equation reads

$$\frac{\partial p(\ell,t)}{\partial t} = -\frac{\partial}{\partial \ell}\left[\left(Je^{-\ell} - J - \frac{\sigma^2}{2}\right)p(\ell,t)\right] + \frac{\sigma^2}{2}\frac{\partial^2 p(\ell,t)}{\partial \ell^2}, \quad (3.73)$$

and to find the stationary state for $\ell \gg 1$ I set the left-hand side to 0 and take $Je^{-\ell} \approx 0$ to find a differential equation

$$0 = \left(J + \frac{\sigma^2}{2}\right)p'(\ell) + \frac{\sigma^2}{2}p''(\ell), \quad (3.74)$$

that solves as $p(\ell) \propto e^{-\mu\ell}$.

In this case, one sees that the average of the distribution is always well-defined, as $\mu > 1$. This justifies my first assumption, $\overline{W}(t) = \mathbb{E}[W_i(t)]$ for large N and my subsequent study of a stationary state within that assumption. I would also note, as argued by the authors in [64] and in [247], that this corresponds to an instance of the so-called Kesten process [151].

3.3.7 *On the dynamics of inequality*

In the past sections I have focused in finding power-law distributions by finding the stationary distribution of the associated models. This means in particular that I have neglected the study of the dynamics of that convergence.

Take for example either of the models of multiplicative growth in the previous sub-sections. The exponent depends, for example, on the volatility σ^2. What would happen if the system is in its stationary state and then one changes that parameter, so that the new stationary distribution has a different exponent?

This goes beyond a mere exercise: looking at Figure 2.4 it is evident that the effective exponent describing income distribution in the United States has fluctuated quite a lot in time, becoming smaller in the last 40 years. In Ref. [121], the authors study precisely this situation.

It is indeed possible, through the same techniques used later in Chapter 7, to study the dynamic convergence of the models in Sections 3.3.4 and 3.3.5 by transforming the non-hermitian Fokker-Planck evolution into a hermitian Schrödinger's equation. This is technique is described in Appendix B.2.2.

The main finding is that for the two models, i. e. including both a barrier for small values of x and death at a rate λ, the distribution converges to the stationary state as $e^{-t/\tau}$, with

$$\frac{1}{\tau} = \frac{m^2}{2\sigma^2}\mathbf{1}_{m<0} + \lambda, \tag{3.75}$$

and that this convergence is too slow to explain the variations of the exponent in Figure 2.4. Note that the Zipf's law $\mu \approx 1$ corresponds in this case to $\tau \to \infty$, and so converging to a distribution with $\mu = 1$ takes an infinite amount of time! Note that this is also the case of the continuous SSR model in Section 3.1.3.

Gabaix et al. show in [121] that the bare multiplicative growth model is unable to replicate the fast increase in inequality observed empirically since the 1980s. Their answer to this is to augment the models, by introducing for example two types of agents, with a different drift m and volatility σ^2, or by introducing "jumps", where the income or wealth of agents changes from time to time by a significantly larger magnitude in a time-step than allowed by the standard gaussian models. The best explanation, according to their model, of the recent fast increase in inequalities is that the average growth m and volatility σ^2 are not the same for all individuals. Accounting for this allows to find faster transition speeds than the base-line model of multiplicative growth with frictions, and has obvious consequences in the understanding of the origins of wealth and income inequality.

This reminds us that these simple models should be used as a qualitative guide in the understanding of the phenomena they describe, and that reality is often more complicated. It has been reported, for example, that standard multiplicative growth models do not correctly account for the observed dynamical properties that govern city growth [265].

3.4 CRITICAL MODELS FOR POWER-LAWS

Many models showed in the previous section have the particularity of generating exponents that can be tuned within a certain range by picking

good parameters. This was not the case for the "greedy" model presented in Section 3.2, nor for the preferential attachment model presented in Section 3.3.2.1, although for this last model I showed that it could be seen as a specific instance of the Simon-Yule model family.

In this section I will instead explore different models that all reduce to the Poisson branching process. This process has the particularity of having a *critical point*, characterized by a power-law distribution with an exponent $\mu = 1/2$. As with the generic scaling relation introduced at the beginning of Chapter 1 linking the volume of an object with its characteristic length, this exponent is "universal" and does not depend on the underlying details of the model as long as the basic mechanisms are present, meaning that the value $\mu = 1/2$ has a generic physical origin. This aspect of scale invariance near a critical point, that translates into power-law scaling relations with "universal" exponents, is something that can be found in many models in statistical physics that exhibit a *second-order phase transition*.

3.4.1 *Branching processes and the Galton-Watson model*

The following *branching process* is very well known under the name of the Galton-Watson process, as it was postulated by Galton and Watson in 1873 and then solved in 1875 as a model of how some surnames may disappear, and of how many people carry a given surname [268]. It is however argued in [70] that it was first proposed by Bienaymé in 1845, and then by Cournot in 1847 in a different context, which I shall also explicit below.

In the family/surnames context of Galton-Watson and Bienaymé, the model imagines that an individual may reproduce and have and offspring of k individuals with probability φ_k, each of them having then their own progeny. The question then is to determine the probability that that "family" goes extinct after n generations, and also the number of individuals within that family. The Cournot interpretation imagines instead bets: imagine a game where you can buy a ticket costing one coin, where your bet can give you k coins with a probability φ_k, which you then use to buy k tickets to bet again. How many iterations of the game can you play before you run out of money (i. e. when all the bets you do at some point are losing bets)? And how much money will you have played up to that point?

In both cases, the problem can be represented by a tree, as depicted in Figure 3.6. This representation shows the conceptual similarity to the Yule-Simon model. In that model, all the nodes in the tree could reproduce at a rate g, or a new tree could be created with probability s. In this model, only the "leaves" of the tree, meaning the offspring from the last generation in the model, may reproduce, and there is no possibility to start a new tree.

To tackle this problem, I call $p_t(n)$ the probability that the t^{th} iteration has n descendants. The corresponding random variable is denoted N_t. Quite straightforwardly, I can write that

$$p_{t+1}(n) = \sum_m p_t(m) \sum_{n_1,\dots,n_m} \varphi_{n_1} \dots \varphi_{n_m} \delta_{n_1+\dots+n_m,n}, \tag{3.76}$$

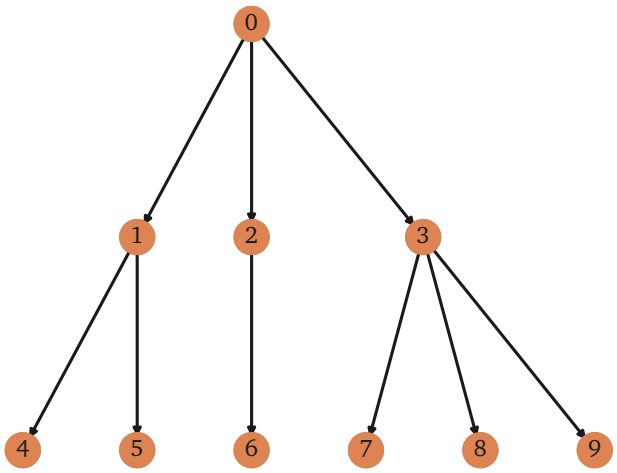


Figure 3.6: A realization of the branching process for three generations, or three betting iterations. Here, the node labelled 0 had an offspring of 3 nodes: 1, 2 and 3. Each of them then led to a different number of offshoots. In the Galton-Watson interpretation, 0 is a patriarch, and here the family has 10 individuals and may or not go extinct at the next iteration. In the Cournot interpretation, the first bet earned the player 3 coins, which he then used to play again, earning 6 coins in the last iteration, which he can use to play again, etc.

meaning that to have n descendants at step $t+1$ you have to sum over all the possibilities of having m descendants at step t with the proper offspring. The key then is to introduce a formal tool called a generating function, namely

$$\widehat{p}_t(z) := \sum_n p_t(n) e^{-zn}. \tag{3.77}$$

For example, $\widehat{p}_1(z) = \sum_n \varphi_n e^{-zn}$. Multiplying Eq. (3.76) by e^{-zn} and then summing over n leads to

$$\begin{aligned} \sum_n p_{t+1}(n) e^{-zn} &= \sum_m p_t(m) \sum_n \sum_{n_1,\ldots,n_m} \varphi_{n_1} \cdots \varphi_{n_m} e^{-zn} \delta_{n_1+\ldots+n_m,n} \\ &= \sum_m p_t(m) \widehat{\varphi}(z)^m, \end{aligned} \tag{3.78}$$

where $\widehat{\varphi}(z) = \sum_n \varphi_n e^{-zn}$.

Letting now $f(z) = -\ln(\widehat{\varphi}(z))$, the recurrence equation reads

$$\widehat{p}_t(z) = \widehat{p}_0(f^t(z)), \tag{3.79}$$

where the superscript $f^t(z)$ stands for the function f composed t times.

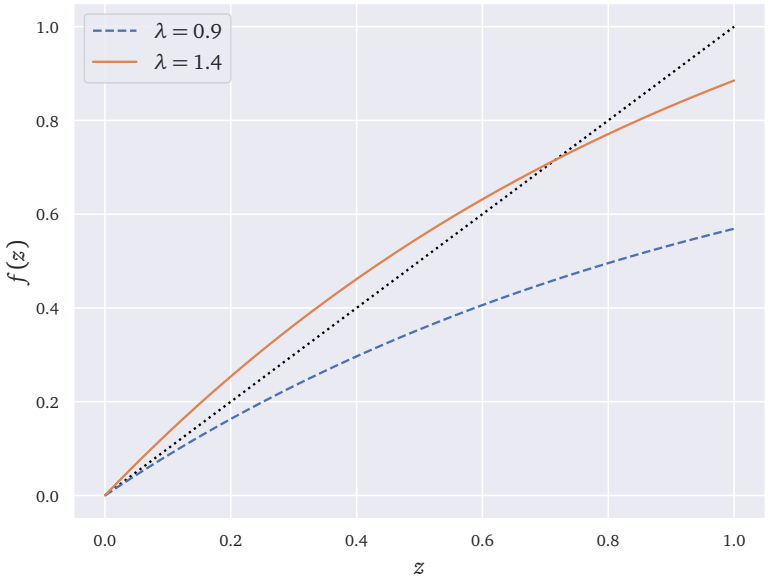


Figure 3.7: Fixed point equation for the Poisson branching process. Note that when $\lambda < 1$ there is a fixed point $z^\star > 0$, while for $\lambda > 1$ the only fixed point is given by $z^\star = 0$.

We can also find the average number of descendants at the t^{th} generation by differentiating Eq. (3.79). This number, denoted $\mathbb{E}[N_t]$, satisfies

$$\mathbb{E}[N_t] = \overline{k}^t, \tag{3.80}$$

where $\overline{k} = f'(0) = \sum_k k\varphi_k$ is the average number of descendants per individual. The reader can therefore already guess that if $f'(0) < 0$ the family is likely to go extinct.

In fact, the probability $p_t(0)$ that the family is extinct, or that the player has lost all his money, at the t^{th} generation is given by $\widehat{p}_t(z)$ for large z. Given the form of Eq. (3.79) this limit can be understood by studying the recursion equation $z_{t+1} = f(z_t)$, and therefore by studying the fixed points of f.

This can be studied on its own for any distribution φ_k. A realistic setting is instead given by the situation where the number of descendants follows a Poisson distribution of parameter $\overline{k} = \lambda$,

$$\varphi_k = \frac{\lambda^k e^{-\lambda}}{k!}, \quad f(z) = \lambda\left(1 - e^{-z}\right). \tag{3.81}$$

The function $f(z)$ is plotted in Figure 3.7, and one sees that there are two fixed points for $\lambda > 1$: an unstable fixed point at $z^\star = 0$ and a second stable fixed point $z^\star > 0$. In this case $\lambda > 1$, it means that

$$\lim_{t\to\infty} \widehat{p}_t(z) = e^{-z^\star}, \tag{3.82}$$

and since $p_t(0) = \lim_{z\to\infty} \widehat{p}(z)$, the probability that the family survives is $1-e^{-z^\star} < 1$. When $\lambda < 1$ the only fixed point is $z^\star = 0$, and so $p_t(0) \to 1$ for large t: the family is almost surely extinct at long times.

Note that here $\lambda_c = 1$ is a *critical point*, a value of the parameter λ that separates two wildly different macroscopic behaviours, just as the points $(0°, 1013\text{hPa})$ and $(100°, 1013\text{hPa})$ are critical points separating the solid, liquid and gaseous states of water.

3.4.1.1 *Family size distribution*

The previous section studied the size N_t of the t^{th} generation. It's also possible to study the size of the whole family, the quantity $S = 1 + \sum_t N_t$. Let $P(s)$ be the probability that such a family is of size s, then it is evident that

$$P(s) = \sum_k \varphi_k \sum_{s_1,\ldots,s_k} P(s_1)\ldots P(s_k)\delta_{1+s_1,\ldots,s_k,s}, \tag{3.83}$$

as it is determined by the probability of having k descendants, each with an offspring s_k, the sum of which is s.

As before, I introduce a similar generating function,

$$g(x) = \sum_s P(s)x^s, \tag{3.84}$$

which one can show satisfies

$$g(x) = x\sum_k \varphi_k g(x)^k = x\widehat{\varphi}\left(-\ln\left(g(x)\right)\right). \tag{3.85}$$

In the case where φ_k describes a Poisson distribution of parameter λ, this implies

$$g(x) = xe^{\lambda(g(x)-1)}, \tag{3.86}$$

which, after a bit of algebra, allows one to find an explicit expression for $P(s)$ [206], as

$$P(s) = \frac{(s\lambda)^{s-1}}{s!}e^{-s\lambda}, \tag{3.87}$$

but note that a general expression can be obtained in terms of the derivatives of $\widehat{\varphi}$.

In the large s limit, one can use Stirling's formula $s! \approx \sqrt{2\pi s}\left(\frac{s}{e}\right)^s$, to find

$$P(s) \approx \frac{e^{-s(\lambda-1-\ln(\lambda))}}{\lambda\sqrt{2\pi}}\frac{1}{s^{3/2}}, \tag{3.88}$$

that becomes a pure power-law $P(s) \propto s^{-1-\mu}$ with $\mu = 1/2$ in the limit $\lambda \to 1$.

3.4.1.2 *Erdös-Renyi graphs*

As hinted by the tree-like representation in Figure 3.7, the previous model can also be used to study networks that are "tree-like" in a sense that I will define below. This is the case of an Erdös-Renyi random network [107]: consider a network of N nodes, where there is an edge $i \longleftrightarrow j$ with probability $p = \frac{\lambda}{N}$. The probability φ_k that node i has k neighbours is thus given by

$$\varphi_k = \binom{N}{k}\left(1-\frac{\lambda}{N}\right)^{N-k}\left(\frac{\lambda}{N}\right)^k \approx \frac{\lambda^k}{k!}e^{-\lambda} \tag{3.89}$$

for large N. The parameter λ is therefore the average degree of each node.

Furthermore, the probability to have a loop of length l, denoted $i_0 \to i_1 \to \ldots \to i_{l-1} \to i_0$, is $\left(\frac{\lambda}{N}\right)^l$ because of independence, and goes to 0 for $N \gg 1$. This is what is meant by saying that the graph is "tree-like", as it locally resembles the tree in Figure 3.7.

The ideas we've used above can then be used to study the distribution of the sizes of clusters in this graph, as the Galton-Watson generative process can be transposed into a graph discovery algorithm that runs as follows. Consider indeed that you are exploring the graph starting from a randomly chosen node i: to find the size of the cluster it belongs to, you simply go through the neighbours of i, then their neighbours and so on. Since there are no loops, each neighbour in the graph is the analogue of a child event in the Galton-Watson process, and once all nodes in the cluster have been accounted for it is as if the family perished. This analogy, to my knowledge, was first highlighted in the context of polymer physics to explain their size distribution [250].

The size of the cluster is therefore the exact equivalent of the total size of the family S in the previous section, and it follows that clusters are distributed according to Eq. (3.87). If one takes a given instance of an Erdös-Renyi network, then one can look at the probability $\tilde{P}(s)$ that a randomly chosen cluster has a size s. This should verify

$$P(s) = \frac{s\tilde{P}(s)}{\sum_{s'} s'\tilde{P}(s')}, \tag{3.90}$$

and so $\tilde{P}(s) \propto P(s)/s$, and the corresponding power-law exponent should be instead $\mu = 3/2$ when close to the critical value $\lambda_c = 1$.

Note that here I have shown a scenario where a power-law distribution emerges in the sizes of the connected components in a network, but the degree distribution here is poissonian.

3.4.2 *The Random Field Ising model and avalanches*

The next model was first proposed as a way of understanding anomalous properties of disordered magnets [235, 236], and in particular for the observed hysteresis in their response to an external field. Placing these objects

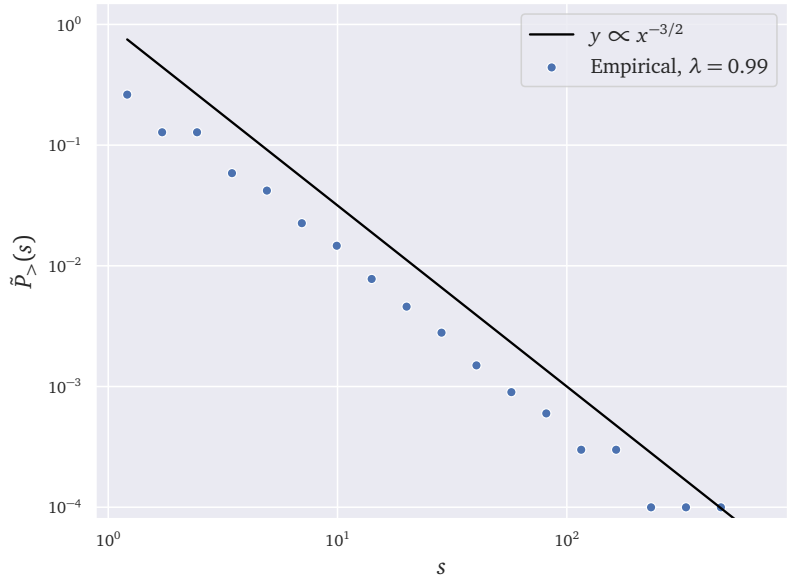


Figure 3.8: Complementary cdf of the distribution of sizes of clusters in an Erdös-Renyi random network. This corresponds to an instance of the graph with $2 \cdot 10^4$ nodes, each having an average degree of $\lambda = 0.99$. The sizes of the clusters are distributed according to $\tilde{P}(s) \sim s^{-1-\mu}$, with $\mu = 3/2$.

in an external magnetic field causes the individual spins in the magnet to align with it, but as one slowly varies the external field the individual spins "flip" in intermittent bursts or avalanches, with a characteristic signature that is known as *Barkhausen noise*.

The model can also be transposed to model binary decision situations [122, 135, 196, 185, 63], where agents have initial preferences for either choice, but also a global incentive to take one choice or another and a tendency to imitate their peers. It consists then of a system of N agents that make a binary choice, represented by a variable $S_i = \pm 1$. There are different variants of the model, but for clarity I choose to think of the agents as maximizing a certain "utility" function, which for agent i reads

$$U_i = S_i \left(h_i + F + \sum_j J_{ij} S_j \right). \tag{3.91}$$

The interpretation of the different terms in Eq. (3.91) is simple: h_i is a *"local field"* encoding the agent's idiosyncratic preference for either choice, and for example large positive values of h_i bias the decision towards $S_i = +1$; F is the *"external field"* and describes global incentives and information that are the same for everyone, and finally the last term describes that an agent's decision depends on what his peers do, through what is known in models for ferromagnetism as the *"exchange interaction"* J_{ij}. For example, if J_{ij} is large and positive this will bias the decision towards $S_i = S_j$,

describing imitation, while large, negative values of J_{ij} do the opposite and bias towards $S_i = -S_j$. For simplicity, I consider the "mean-field" version of the model, and take $J_{ij} = J/N$ with $J > 0$, so that the utility reads

$$U_i = S_i \left(h_i + F + Jm\right), \tag{3.92}$$

with

$$m = \frac{1}{N}\sum_i S_i \tag{3.93}$$

the "magnetization", so called because of the initial physical interpretation of the model, a quantity representing the aggregate choice made by individuals.

One can easily gain intuition by considering extreme cases. When $J = F = 0$, the agent's decision is very simple, and given by $S_i = \text{Sign}(h_i)$, and so all agents make different choices. When $F \to \pm\infty$ all agents make the same choice, $S_i = \text{Sign}(F)$. However, when J is very large the opinion of all the agents is determined self-consistently, as $S_i = \text{Sign}\,(m)$: each agent aligns herself with the majority, but this majority is itself determined by the individual actions of all agents!

This self-consistency shows the way to solve the model. For each agent the optimal decision is given by

$$S_i = \text{Sign}\left(h_i + F + Jm\right), \tag{3.94}$$

and this equation may now be summed over and divided by N, to find the self-consistent equation that m should satisfy,

$$m = \frac{1}{N}\sum_i \text{Sign}\left(h_i + F + Jm\right) \approx \mathbb{E}\left[\text{Sign}\left(h + F + Jm\right)\right], \tag{3.95}$$

since when N is large, one can model the h_i fields with random variables, and the right-hand side becomes an average over them. Introducing their density $\rho(h)$, this equation becomes

$$m = 2\int_{-\infty}^{Jm+F} \mathrm{d}h\ \rho(h) - 1 := R(Jm + F), \tag{3.96}$$

and it therefore becomes a fixed-point equation for the function R. For clarity and ease of computation I take a Laplace distribution for ρ, namely

$$\rho(h) = e^{-|h|/\Delta}/(2\Delta), \tag{3.97}$$

but the results I am presenting are general and do not depend on the precise form of the distribution, as long as it's smooth. For $F = 0$, the number of solutions to the fixed point equation $m = R(Jm)$ depends on the slope of the tangent at 0, given by $2J\rho\,(0)$. The analysis below is restricted to the simple case where ρ is even and mono-modal. If $2\rho\,(0)$ is larger than 1 then there are three fixed points, $(-m^\star, 0, m^\star)$, with 0 an unstable fixed

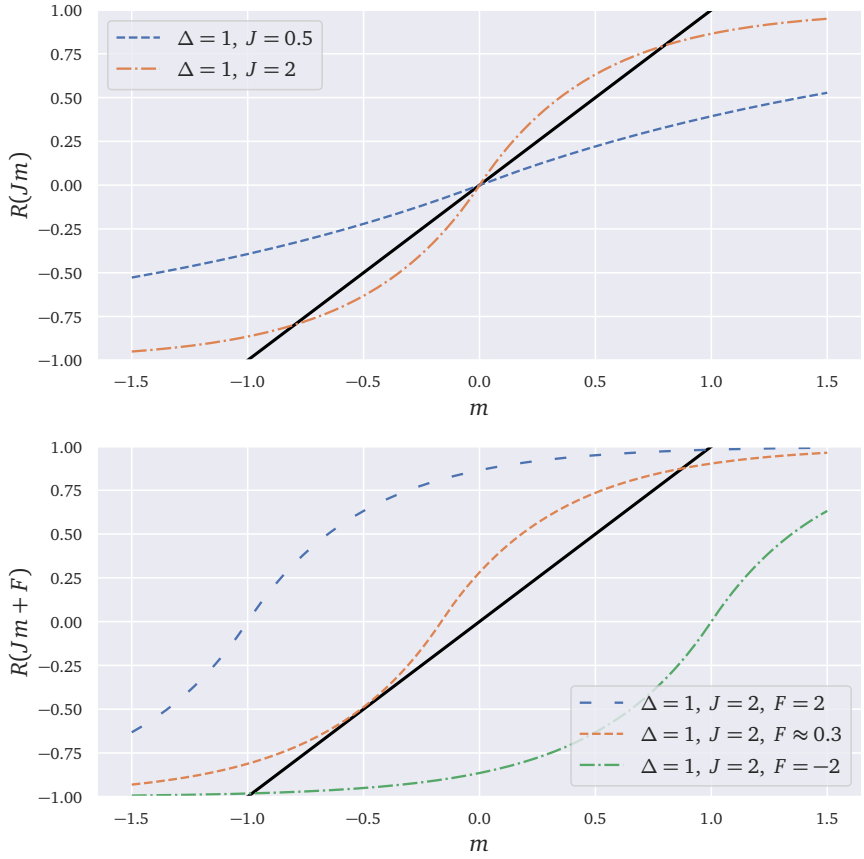


Figure 3.9: Graphical solution to the fixed point equation $m = R(Jm + F)$. The density ρ is that of a Laplace distribution, with $\Delta = 1$ for simplicity. The top panel corresponds to $F = 0$, and there is a transition at $J = 1$, between a setting where the equation has 3 different solutions to a solution where there is none. In the bottom panel, as F is varied the number of solutions can change from just one to three if $J > 1$; here the middle curve sits right at the critical point, where the system can go from having 1 to up to 3 solutions.

point to the iterative sequence $m_{t+1} = R(Jm_t)$; when it is smaller than 1 there is a single solution at 0. Note that in the general case, for example when ρ has more than one mode, a larger number of solutions can appear.

In terms of the Laplace distribution, it is direct that the critical point is determined by

$$J = \Delta. \tag{3.98}$$

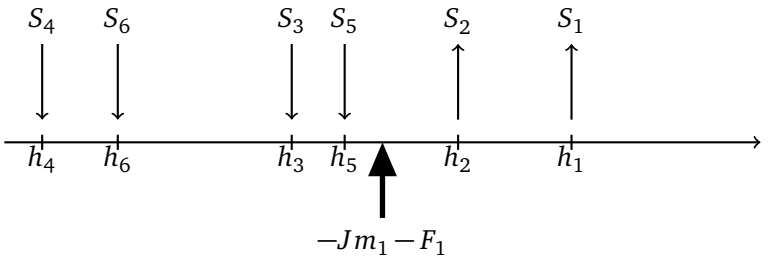
A graphical illustration of this is given in Figure 3.9. The interpretation is clear, when there is a strong enough interaction term $J > \Delta$, one can imagine a setting where one starts from an initial value $F \gg J$. In this case, the S-shaped curve in the Figure is completely shifted to the left, and the only solution is $m \approx 1$. Decreasing F slowly shifts the curve to the right, until it intersects the curve at more than one point. The slope of the

tangent of the S curve at that critical point should be 1, and therefore the full condition that these critical points satisfy is

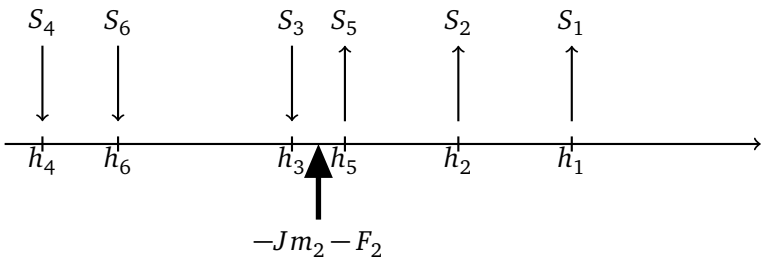
$$1 = 2\rho\left(Jm + F\right). \tag{3.99}$$

The interpretation to this phenomenology is clear: if social pressure is too strong, there may be situations where a choice persists exclusively because of peer imitation, despite a weak "global incentive". In particular, if one starts with $F \gg J$, the initial choice is $S_i = 1$ for nearly every agent in the model. The value of F can then become negative, but it's possible that the average choice is still $m > 0$, and so a majority still make the choice $S_i = 1$. As F goes below a threshold where the three fixed points disappear, the system suddenly changes to $m < 0$, and therefore the majority of agents change their mind from $S_i = 1$ to $S_i = -1$. Interestingly, if the experiment is run in the opposite direction, that is from negative to positive F, the system does not behave in the same way, since the initial situation $m < 0$ persists for small but positive F. This explains the hysteretic behaviour of the model.

Yet the question to be asked is: What is the size distribution of the "avalanches" of opinion shifts in this model? To assist in this, let us introduce the following graphical representation for a system with a finite number of agents. I take for simplicity 6 agents, the choice of each represented with an arrow, so that $S_i = 1 :=\uparrow$ and $S_i = -1 =\downarrow$, with a given initial value for F:



with all $S_i = 1$ if $h_i > -m_1 - F_1$ and $S_i = -1$ in the opposite case. You can then imagine decreasing the value of F slightly, to a new value F_2. This may then cause the site labelled 5 to flip, and therefore there will also be a new value for the average m_2. The new situation can be represented as



and in this case the change was not strong enough to have agent number 3 flip as well. In this case, the difference between the two values of m is given by

$$m_2 = m_1 - \frac{2}{N}, \tag{3.100}$$

and no other agent changed their mind because there wasn't a variable h_i satisfying $-Jm_1 - J\frac{2}{N} - F < h_i < -Jm_1 - F$. There could, however, have been one with a probability

$$p = \frac{2}{N}\rho\left(Jm + F\right), \tag{3.101}$$

in which case the other agent would have changed their mind, changing the value of m again and triggering more opinion changes. Owing to the form of Eq. (3.101), these avalanches are determined by a Poisson branching process in the limit $N \to \infty$. The probability distribution for the size of an avalanche is then given by

$$P(s) = \frac{(s\lambda)^{s-1}}{s!}e^{-s\lambda}, \quad \lambda := 2\rho(m + F), \tag{3.102}$$

as before. And therefore at the critical point defined in Eq. (3.99) the distribution of the sizes of the avalanches is given by a power-law as in Eq. (3.88), with an exponent $\mu = 1/2$.

3.4.3 *Sweeping the critical point*

The downside of the previous models is that they only provide power-laws at the *critical point*, corresponding to when the parameter we've called $\lambda = 1$. Instead, one can imagine that one observes the distribution of sizes corresponding to different branching processes, albeit with different values of λ. To my knowledge, this mechanism has not been proposed elsewhere in the literature, but it seems realistic enough to warrant a comparison with data.[4]

As an example, imagine the following branching process. Imagine a book that has just been published, a first person reads it and manages to persuade

4 Despite preliminary explorations, I have not yet found a data source yielding conclusive enough results to be added to this work.

on average λ people to read it, who then attempt to persuade other people themselves. The parameter λ should naturally be book-dependent, as one would imagine it to be related to its attractiveness. The number of books sold given a certain value for λ should be determined by Eq. (3.87), but to determine the distribution of the sales of many *different* books one should integrate over the different possible values that λ can take.

Hence, I write $\lambda = 1+\varepsilon$, and imagine ε to be distributed with a certain distribution $p_\varepsilon(\varepsilon)$ that is very peaked around 0. The distribution of all sizes is given by integrating the distribution conditional on ε, which for large s would have a first contribution for all values $\varepsilon < 1$ proportional to

$$s^{-3/2}\int_{-\infty}^{0} \mathrm{d}\varepsilon\; p_\varepsilon(\varepsilon)\,\frac{e^{-s(\varepsilon-\ln(1+\varepsilon))}}{1+\varepsilon} \approx s^{-3/2}\int_{-\infty}^{0} \mathrm{d}\varepsilon\; p_\varepsilon(\varepsilon)\, e^{-s\frac{\varepsilon^2}{2}}. \quad (3.103)$$

Substituting $y = \sqrt{s/2}|\varepsilon|$ in the integral gives a term $1/\sqrt{s}$ because of the jacobian of the transformation, times a convergent integral that doesn't depend on s, which I denote $I_0(p_\varepsilon)$. Gathering everything, this contribution to the distribution of avalanche sizes is a power-law, corresponding to a Zipf law, i. e. a contribution proportional to $s^{-3/2}s^{-1/2} = s^{-2}$.

For the contribution corresponding to positive values of ε, I showed in Sec. 3.4.1 that for a given value of $\varepsilon > 0$, there is a finite probability to have a system-wide avalanche, or equivalently that the Poisson branching process spans a number of agents commensurable with the size of the system. This probability is given by

$$\mathbb{P}(S \approx N) = 1 - e^{-z^*}, \quad z^\star = (1+\varepsilon)\left(1-e^{-z^\star}\right), \quad (3.104)$$

this means that the contribution to the density coming from $\varepsilon > 0$ reads

$$\int_0^\infty \mathrm{d}\varepsilon\; p_\varepsilon(\varepsilon)\left(1-e^{-z^\star(\varepsilon)}\right)h(s-N) := I_1(p_\varepsilon)h(s/N), \quad (3.105)$$

where $h(s/N)$ is a "hump" that is very localized at values close to N.

With this, the complete distribution of avalanche sizes for large N and large S is given by

$$P(s) \approx I_0(p_\varepsilon)\,s^{-2} + I_1(p_\varepsilon)\,h(s/N), \quad (3.106)$$

and can be interpreted as mixing both "black swan" events [257], that is events that are extreme because they are power-law distributed, but that are not system-wide, with "dragon kings" [248, 246], events that are so extreme that they don't fall into a power-law classification.

3.4.4 *Self-organized criticality*

Another explanation of why this kind of system may sit close to the critical point is that an external mechanism may drive it close to it. This is called *self-organized criticality*, and was a scenario that was proposed by Bak, Tang

and Wiesenfeld [33, 31] as a way of describing how "Large interactive systems naturally evolve toward a critical state in which a minor event can lead to a catastrophe" [85].

This scenario can be understood simply with an analogy of a sandpile. Imagine a growing sandpile, on top of which you keep adding grains of sand. As long as the slope at the top of the pile is small you can safely keep adding grains, but the pile gets steeper and steeper. At a certain point, the pile is too steep: a single grain of sand can fall-off, and bring other grains down in its fall. This happens as we've kept adding energy to the system, up to the point where the energy it stores is so large that anything can trigger a dissipative event.

The simplified model consists of a cellular automata model defined in a d-dimensional lattice. Each site in the lattice is associated with an integer variable called the energy. Over time, energy is added to each site, and once a site's energy reaches a certain threshold its energy "relaxes" and is transferred to its neighbours. A single relaxation can then trigger other relaxations. In the limit $d \to \infty$, which simply means that a site can transfer its energy to all the other sites in the system,[5] it can be shown that the system becomes equivalent to a Poisson branching process [29, 281], but where the parameter λ follows an evolution of the form

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} \approx 1-\sigma(t), \tag{3.107}$$

where σ is a random term related to the size of the avalanche at time t, with a mean proportional to the average size of an avalanche for a given value of λ, i. e. $\sum_n \lambda_n = (1-\lambda)^{-1}$, and random fluctuations about that value.

Thus, λ keeps growing towards the critical value $\lambda = 1$, but as this happens the avalanches become larger and larger, so that when λ approaches 1 it is immediately pushed back from it. As the reader will see, this very simple toy-model can give very realistic scenarios to explain power-law distributions and anomalous volatility in socio-economic systems.

3.4.5 *Second-order phase transitions*

The Ising model is the best-known example of a physical system characterized by power-laws near its critical point. It consists of binary "spin" variables $s_i = \pm 1$, where i represents a magnetic moment associated to an ion in a lattice (a periodic graph representing a metallic crystal). Each spin

5 Indeed, in a d-dimensional square lattice each site has $2d$ neighbours: when $d = 2$ there is for example a neighbour to the right, to the left, to the top and below. When $d \to \infty$ the number of neighbours simply becomes commensurable with N, the total number of sites in the lattice.

interacts with an external magnetic field h and with its neighbours in the lattice, and the associated Hamiltonian is

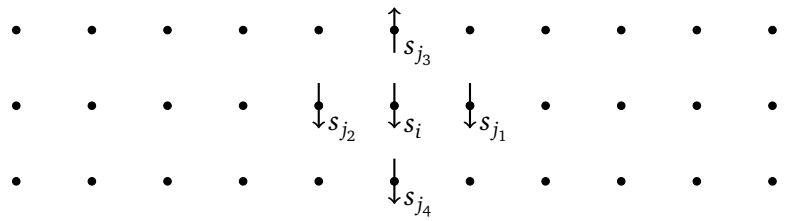
$$\mathscr{H}(\vec{s}) = -J \sum_{\langle i,j \rangle} s_i s_j - hm, \tag{3.108}$$

with $\vec{s} = (s_1, \dots, s_n)$ and $m := \frac{1}{N} \sum_i s_i$ the magnetization. The probability associated to a certain configuration $\vec{s}$ is then given by

$$p(\vec{s}) = \frac{1}{Z} \exp\left(-\beta \mathscr{H}(\vec{s})\right), \quad Z = \sum_{\vec{s}} \exp\left(-\beta \mathscr{H}(\vec{s})\right), \tag{3.109}$$

where $\sum_{\vec{s}}$ means to sum over all possible configurations, i. e. $\sum_{\vec{s}} = \sum_{s_1=\pm 1, \dots s_n = \pm 1}$. The quantity Z is called the *partition function*. The parameter β is the inverse of the temperature, and when $J > 0$ it is a model for ferromagnetic materials, since it means that in absence of an external field a magnetic moment has a tendency to align with its neighbours. For $\beta \to \infty$, or equivalently at 0 temperature, the system is found in its lowest energy state, corresponding to either all $s_i = 1$ or all $s_i = -1$.

As an example, I have represented an Ising configuration for a 2-dimensional square lattice below,



where I only show the spin variables of a certain site i and of its neighbours $j_1, \dots, j_4$, along with just a portion of the lattice. In this schematic representation, one has $s_i = -1 :=\downarrow$ as the most probable value because 3 of its neighbours also point down.

When the dimension of the lattice is $d > 1$ there exists a value $\beta_c < \infty$ above which there is a cooperative phenomenon known as *spontaneous magnetization*, where the magnetization $m \neq 0$ even when the external field $h = 0$ [83]. This means that more than half of the spins are all equal to either -1 or $+1$. For $\beta < \beta_c$, i. e. in the high-temperature regime, roughly half the spins are negative and the other half is positive, and $m = 0$.

For example when the dimension is $d = \infty$, I rescale $J \to J/N$ to get the mean-field Ising Hamiltonian

$$\mathscr{H}(\vec{s}) = -Jm \sum_i s_i, \quad m := \frac{1}{N} \sum_i s_i. \tag{3.110}$$

From here, I can compute the average value $\mathbb{E}[s_i] = m$ using Eq. (3.109) to get

$$m = \tanh(\beta J m), \tag{3.111}$$

which using fixed-points arguments as before has single solution $m = 0$ when $\beta < \beta_c := \frac{1}{J}$, with two additional solutions appearing when $\beta > \beta_c$.

Onsager [205] showed that for $d = 2$ the critical temperature $T_c = 1/\beta_c$ equals

$$T_c = \frac{2J}{\ln\left(1+\sqrt{2}\right)}, \tag{3.112}$$

and for example the magnetization m close to criticality scales as

$$m \underset{T \approx T_c}{\propto} \left(T - T_c\right)^{1/8}, \tag{3.113}$$

hinting already at the scale-free behaviour in the model. This exponent $1/8$ only depends on the dimension of the lattice, and not on its precise arrangement. For example, in a 3-dimensional lattice it has a slightly higher value, and is found to be ≈ 0.313.[6]

In fact, denoting by $\vec{i}$ the position of spin i in the lattice and therefore $\left|\vec{i} - \vec{j}\right|$ the distance between two spins, the spin-spin correlation function defined as

$$C\left(\left|\vec{i} - \vec{j}\right|\right) = \mathbb{E}\left[s_i s_j\right] - \mathbb{E}\left[s_i\right]\mathbb{E}\left[s_j\right], \tag{3.114}$$

decays as a power law at the critical-point [150], i. e.

$$C\left(\left|\vec{i} - \vec{j}\right|\right) \propto \frac{1}{\left|\vec{i} - \vec{j}\right|^{d-2+\eta}}, \tag{3.115}$$

where η is an exponent dependent on the dimension of the model. For $d = 2$ the exact value is $\eta = \frac{1}{4}$, and for $d = 3$ it's $\eta \approx 3.6 \cdot 10^{-2}$. This decay has a geometric interpretation: if one were to look at a 2-dimensional Ising lattice with spins $s = 1$ coloured black and white spins $s = -1$, then the observed figure would be *scale-invariant*, meaning that its overall aspect would be unchanged by zooming in or out. This is depicted on Figure 3.10, where I've simulated a two-dimensional Ising model at the critical temperature given in Eq. (3.112), with dark and clear clusters and a clear fractal structure.

3.5 SUMMARY AND CONCLUSION

As a recapitulation, these are the main mechanisms I've covered throughout this section which give a distribution with a power-law tail $\propto s^{-1-\mu}$:

6 Solving the 3-dimensional Ising model remains one of the most interesting unsolved puzzles in statistical mechanics. It is telling that one of the most promising approaches to a solution exploits the scale-free nature of the model close to criticality, using an approach known as *conformal bootstrap* that also has applications in high-energy physics [106]. For a vulgarized take on this, see [277].

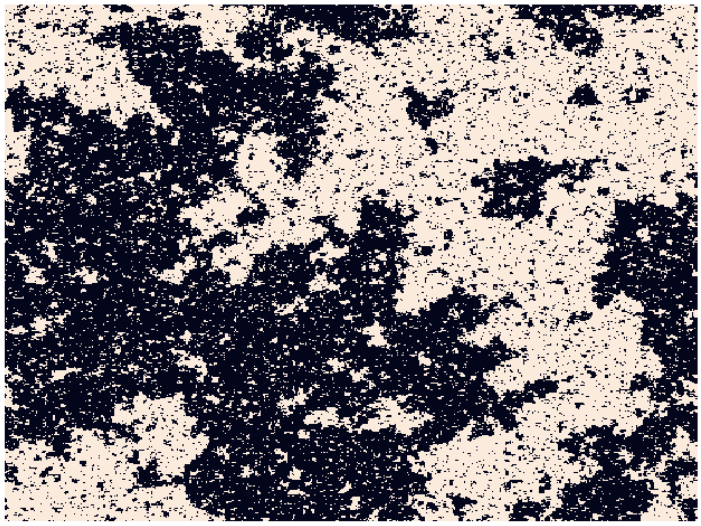


Figure 3.10: Critical 2-dimensional Ising model on a 500×500 lattice with periodic boundary conditions. Notice the *scale-invariance*, with white clusters inside black clusters and so on. This is in line with the power-law behaviour of the spin-spin correlation function $C(\left|\vec{i}-\vec{j}\right|) \propto \left|\vec{i}-\vec{j}\right|^{-1/4}$.

THE MONKEY TYPIST MODEL for which one gets an exponent

$$\mu = \left(1 + \frac{\ln(1-p)}{\ln(n)}\right)^{-1}, \tag{3.116}$$

where p is the probability of hitting the space-bar, and n the size of the alphabet. The Zipf law $\mu = 1$ is retrieved in the limit $p \to 0$.

THE YULE/SIMON MODEL here, the exponent reads

$$\mu = 1 + \frac{g}{s} = \frac{1}{1-\alpha}, \tag{3.117}$$

where g is the rate of appearance of new genera, and s the intra-species mutation rate in the Yule interpretation, and α is the probability to start a new family in the Simon interpretation. The Zipf law $\mu = 1$ is obtained in the limit $\alpha, g \to 0$, as the probability of starting new entities, be they new words/families/genera, goes to 0. Note that the Barabási-Albert preferential attachment model corresponds to $g = s$.

SAMPLE-SPACE REDUCING MODELS lead to a power-law distribution regardless of the underlying "energy" distribution $\mathscr{P}(E)$, with an exponent

$$\mu = 1 + \phi, \tag{3.118}$$

a Zipf law in the limit $\phi \to 0$.

The "greedy" version of the SSR model gives instead an exponent $\mu = 1/2$, again regardless of the distribution $\mathscr{P}(E)$, with an exponential cut-off that scales as $(\phi)^{-1}$.

THE RANDOM ENERGY MODEL that consists of N log-normal variables $x_i \propto e^{\sigma\sqrt{t}\xi_i}$ with gaussian ξ_i. It's equivalent to a power-law distribution with parameter

$$\mu = \sqrt{\frac{2\ln(N)}{\sigma^2 t}}, \tag{3.119}$$

in the large T, N limit with μ fixed.

MULTIPLICATIVE GROWTH WITH DEATH AND/OR BARRIER a geometric brownian motion with log-drift m and log-volatility σ^2, supplemented with a death rate $\lambda \geq 0$. An exact power-law is found in the stationary state when $\lambda > 0$, with an exponent

$$\mu = -\frac{m}{\sigma^2} + \frac{\sqrt{m + 2\lambda\sigma^2}}{\sigma^2}. \tag{3.120}$$

Considering instead a model *without death*, i. e. with $\lambda = 0$, but with a lower threshold under which the variable $x(t)$ may not go, leads also to an exact power law when $m < 0$, with an exponent

$$\mu = \frac{|m|}{\sigma^2}. \tag{3.121}$$

THE BOUCHAUD-MÉZARD MODEL very similar in spirit to the multiplicative growth with reflecting barrier model. The exchange/redistribution term J has a similar role to the reflective barrier defined above, and one finds that the (rescaled) variable is asymptotically distributed as a power law, with exponent

$$\mu = 1 + \frac{2J}{\sigma^2}, \tag{3.122}$$

yielding a Zipf law in the limit $J \to 0$.

POISSON BRANCHING PROCESSES AND AVALANCHES these models always give an exponent

$$\mu = 3/2 \tag{3.123}$$

at their *critical point* regardless of the underlying details (i. e. for Galton-Watson, Cournot/Bienaymé, Erdös-Renyi graphs, RFIM...). This is similar to what happens in models exhibiting a second order phase transition in statistical mechanics, as with the Ising model in dimension $d > 1$. The

superposition of many instances of this process with values λ close to its critical value $\lambda_c = 1$ leads to the superposition of a Zipf law with $\mu = 1$ with "system-wide" avalanches or clusters.

Finally, looking back at the objective set in Chapter 1 I have effectively shown that power-laws are found in many complex systems, and that there are common, over-arching mechanisms that explain them. The present review is naturally incomplete, but should give a satisfactory overview of the main facts around power laws and the mechanisms that can engender them. Concerning their consequences, in addition to those surrounding the societal issue of wealth and income inequality, the next Section will discuss their role in large economic fluctuations.

Part II

LARGE MACROECONOMIC FLUCTUATIONS

Why is the economy so volatile? How do small "shocks" to small entities, be they firms or individuals, evolve into large macroeconomic fluctuations? In this section I explore different clues to this decade-old mystery.

The reader will become familiar with the "granularity" argument, where the fat-tailed nature of the economy allows to explain for some of these fluctuations. I will then delve into models that attempt to explain these fluctuations as coming from shocks that are propagated and amplified by the input-output network linking firms, with an original proposition of a realistic scenario for *self-organized criticality* in this context, in a situation similar to models that explore the stability of large ecosystems.

4

CONTEXT: THE BUSINESS CYCLE AND LARGE FLUCTUATIONS

A decades-old puzzle in economics is the persistence of large deviations from the trend in economic time series, such as those describing the Gross Domestic Product (GDP) of countries. A very clear example of this is given in Figure 4.1, where I show the rolling yearly growth rate of the US Real GDP since 1948. The graph shows very wild fluctuations, some of which can be clearly pinned on large "crises", such as the oil shock of 1973, the global financial crisis of $2007-2008$ and, more recently, the unfolding[1] Covid-19 crisis of 2020. In all, the average growth rate in that period is found to be of 3.18%, but the standard deviation is comparable and equals 2.57%.

A natural explanation to this, which in hindsight seems very naive, was proposed in the early 19^{th} century by Herschel and Carrington, but perhaps most famously by Jevons in 1875 (see Ref. [124] and references therein) in what is now known as the "sunspot theory". In short, they proposed that fluctuations of the weather, with Jevons notoriously arguing in favour of those resulting from solar activity, could be the origin of fluctuations in crops' harvests and prices, thereby impacting the whole economy. They were met with scepticism at the time, but their explanation remains present in the idea of exogenous fluctuations.

Within that framework, the economy is thought to be intrinsically in a situation of equilibrium, where all the flows of money and goods should balance each other; the only possible origin of fluctuations then is the "hand of God", i. e. completely exogenous shocks like political crises or natural disasters that temporarily drive the economy out of equilibrium, before having it naturally come back to it through regulatory mechanisms. This in fact lies at the heart of the tools used by central banks to set their policy. Indeed, the "Dynamic Stochastic General Equilibrium" (DSGE) models think of an economy as being made of a "representative firm" that encapsulates all of the industry and a "representative household" that captures the behaviour of labour and consumers, nudged around by random shocks [123]. These were under heavy criticism following the $2007-2008$ financial crisis for their failure to account for large volatility and for the way they broke down when confronted to unusual situations [49, 252].

The causes of large economic fluctuations are therefore of crucial importance, and clarifying their origin would have direct repercussions in public policy. Yet under the framework explained above, that of "representative agents/firms" enduring shocks, it is natural to ask how these idiosyncratic

1 As of the writing of this manuscript.

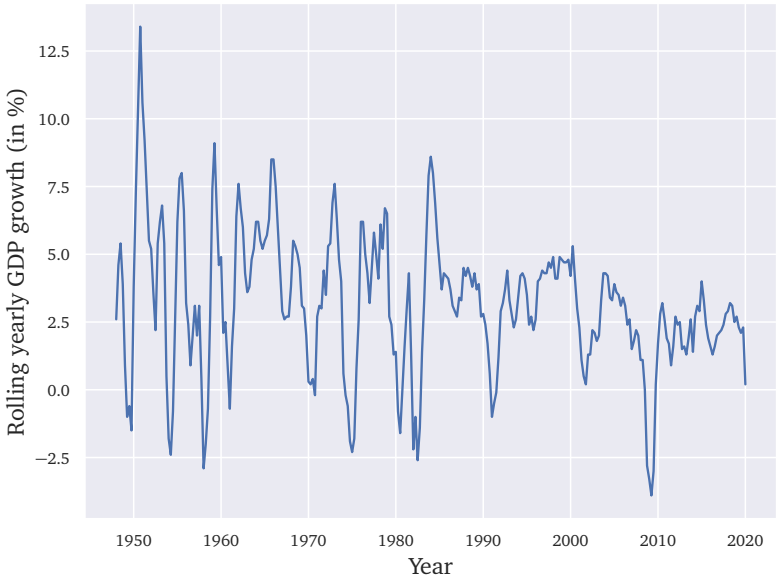


Figure 4.1: Rolling yearly growth rate of the US Real GDP, seasonally adjusted and detrended. The percentage is computed with the variation from the same quarter one year ago. The average since 1948 is of 3.18%, but with a standard deviation of 2.57%.

fluctuations behave at the aggregate level, where they could be amplified or supplemented with fluctuations that are of *endogenous* origin.

Indeed, taking the example of the US economy in the second half of the 20$^{\text{th}}$ century, you could naively imagine it to be made up of N equivalent firms, or N copies of the proverbial "representative firm", each of them making at time t an output $s_i(t) = s(t)$, measured by their sales for definiteness. Their aggregate sales would then be given by

$$S(t) = Ns(t), \tag{4.1}$$

and you could then imagine the sales of each firm to evolve randomly from one year to the next, as

$$s_i(t+1) = (1 + m + \sigma\eta_i(t))\, s_i(t), \tag{4.2}$$

where η_i is a random gaussian variable of unit variance. From one year to the next, one would therefore expect each firm to grow on average by a percentage given by m, with random fluctuations of magnitude σ.

Looking then at the aggregate growth, one should expect to find

$$\frac{S(t+1)}{S(t)} = \frac{\sum_{i=1}^{N}(1 + m + \sigma\eta_i(t))}{N} \approx 1 + m + \frac{\sigma}{\sqrt{N}}\eta, \tag{4.3}$$

with η a gaussian random variable. This amounts to a naive application of the central limit theorem, where the fluctuations about the trend $1 + m$

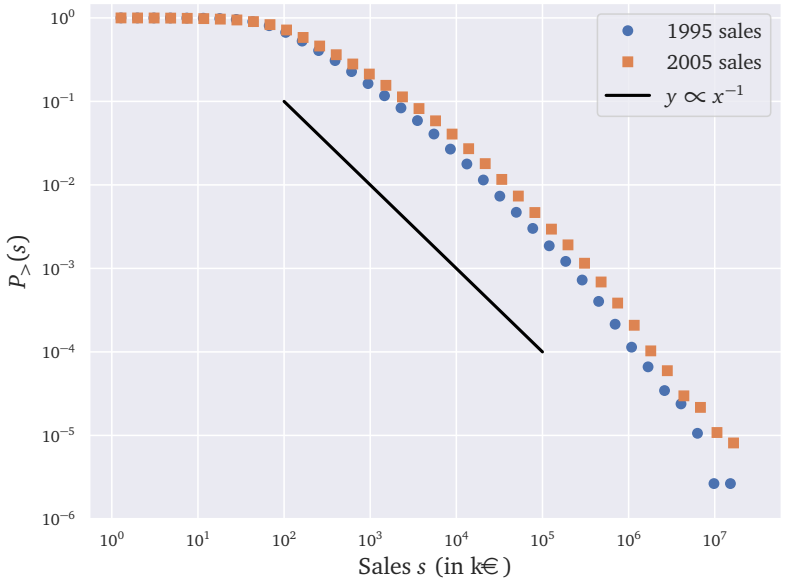


Figure 4.2: Complimentary cdf for the Net Sales of French firms in 1995 and 2005 (non deflated). The source is the INSEE Ficus dataset, filtered down from an initial pool of around 6 million firms identified by their SIREN number. I have selected only those that were present every year from 1994 to 2007 and that declared having at least one employee during that period, amounting in all to 374, 714.

should scale down as $1/\sqrt{N}$, and by this argument one would expect that an economy as large as the United States, counting millions of firms, should not fluctuate at all!

To solve this puzzle, a number of explanations has been put forth, and they consist chiefly of going beyond the hypotheses used above in the application of the central limit theorem. The first consists of dropping the assumption of having "comparable" entities, and is known as the *granularity* argument proposed by Gabaix [120], that I present below.

4.1 GRANULARITY

The starting point of Gabaix' argument is that many quantities in an economics and finance are power-law distributed [119]. As I've showed in Chapter 1, these distributions are indeed ubiquitous. Furthermore, it seems in general that the Zipf law, for which the power-law exponent $\mu \approx 1$, holds for many of these quantities, regardless of the country or time-period under consideration.

To further illustrate this, I have computed the complimentary cdf corresponding to the net sales of a large panel of French firms and of publicly listed NASDAQ/NYSE firms, as well as the complimentary cdf of the market

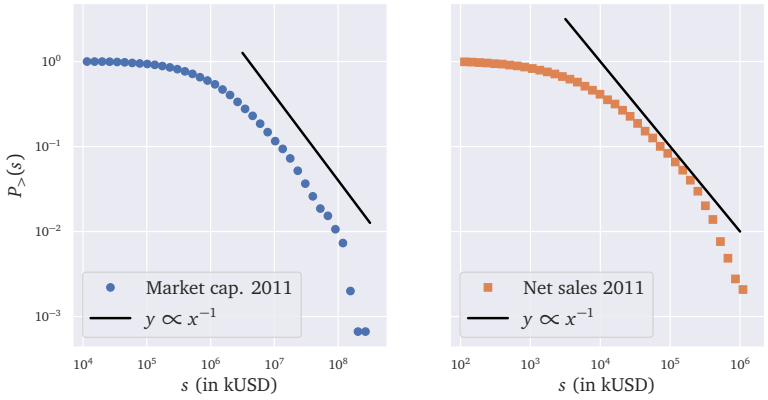


Figure 4.3: Complementary cdfs for the Net Sales and Market Capitalization of publicly traded firms in the NASDAQ/NYSE exchanges. The source is a merged dataset of CRSP/Compustat, totalling 20, 140 firms from 1962 to 2013.

capitalization of the latter. The results are visible on Figure 4.2 for the French INSEE[2] and on Figure 4.3 for the CRSP/Compustat data.

The granularity hypothesis then states that some of the anomalously high volatility in the economy can be directly imputed to large firms. As an example, he cites an OECD study [203] where it is shown that Nokia alone contributed 1.6 percentage points of Finland's GDP growth in 2004. Aggregate fluctuations cannot therefore be thought of as coming from a "continuum" of firms of equivalent size, as the largest firms in an economy constitute incompressible "grains" that are responsible for a large proportion of aggregate volatility. In short, because of the power-law nature of firm sizes the "standard" central limit theorem breaks down, and gaussian intuition must be shed in favour of the usage of Lévy-stable distributions.

As made explicit in Sec. 2.3, the aggregation done in Eq. (4.3) should have random fluctuations that scale down as $N^{\frac{1}{\mu}-1}$ where μ is the exponent of the power-law, much slower than the $1/\sqrt{N}$ decay one would expect in the gaussian case.[3]

For example, considering still that firms grow randomly as in Eq. (4.2), it is possible to write the aggregate sales growth as

$$\frac{S(t+1)}{S(t)} = \sum_{i=1}^{N} (1 + m + \sigma\eta_i) \frac{s_i(t)}{S(t)}, \tag{4.4}$$

2 *Institut national de la statistique et des études économiques*, French national institute for statistics and economic studies.

3 In the case $\mu = 1$, these fluctuations go down as $1/\ln(N)$ instead because of logarithmic corrections.

where $S = \sum_i s_i$ and where I've assumed that all firms grow according to Eq. (4.2) with the same average growth rate m and same volatility σ^2.[4] Taking the variance of Eq. (4.4) it follows that

$$\sigma_S^2 = \sigma^2 H, \quad H = \sum_i \left(\frac{s_i(t)}{S(t)} \right)^2, \tag{4.5}$$

where H is the herfindahl index, also defined in Sec. 2.3.1. When the distribution of the s_i variables is fat-tailed, such that $\mathbb{V}[s_i] = \infty$ or equivalently coming from a power-law distribution with an index $\mu < 2$, one can expect the herfindahl index H to be larger than $1/N$. In this case, the variables $x_i = \left(\frac{s_i(t)}{S(t)} \right)^2$ are power-law distributed with an exponent $\mu/2$ as explained in Sec. 2.3, and their sum H can be written as

$$H \propto N^{\frac{2}{\mu}-2} \xi_{L,\mu/2}, \tag{4.6}$$

with $\xi_{L,\mu/2}$ a Lévy-stable random variable with tail parameter $\mu/2$.

Although more complicated situations are imagined, such as when σ is heterogeneous and depends on the "size" of the firm s_i, this remains the core of the granularity hypothesis: the gaussian intuition, where aggregated shocks average out for a large economy, is not justified because of the fat-tailed distribution of firm sizes. This argument, which I shall revisit in a different context and at a different scale in Chapter 6, was also proposed by Wyart & Bouchaud [279] (see also [254]). Some empirical support has been put forth in favour of this explanation, for example in the original paper [120] but also in [131], but careful empirical exploration has found that large inter-sectoral correlations remain even at a very deep level of disaggregation, and that *network effects* are dominant [114, 2]. This is at odds with the view proposed by the "granularity" hypothesis, where firms endure independent idiosyncratic shocks, and motivates the introduction of production networks as a way of explaining volatility.

4.2 NETWORK MODELS FOR SHOCK PROPAGATION

Considering an economy made of N firms or sectors,[5] a very simplistic way of describing the production π_i of firm i at time t is by writing

$$\pi_i = f_i\big(\{Q_{ij}, \quad j \in \mathcal{N}(i)\}\big), \tag{4.7}$$

where $\mathcal{N}(i)$ is the set of suppliers of firm i and Q_{ij} is the amount of goods produced by j that i holds at the time of production, having bought them in the past. The function f_i is called a *production function* and can be seen

4 This may seem an unrealistic assumption, especially given the results presented later in Chapter 6, but this idea holds in the large N limit as long as the distributions of m or σ have a finite variance.

5 Although in the model presented below the authors speak of *sectors*, I shall speak of firms for definiteness and to keep a similar vocabulary as the one used in Chapter 5

as a phenomenological way of capturing the production technology of i. This is a very "metabolic" interpretation for the role of a firm, as an entity that takes a certain set of inputs and spits out an output with a certain efficiency.

To set ideas straight, i can be pictured as a car-maker and $\mathcal{N}(i)$ is the set of firms j and households that sell the labour and inputs necessary to the production of a car, i. e. tyres, wind-shields, nuts, bolts, etc., but importantly also man-hours. The production function f_i tells us how many cars can be made with a certain amount of inputs Q_{ij}.

The reasoning then is that firms are intertwined, and that a drop in the output of a given firm can wreak havoc and be amplified down the supply chain by disrupting the production of its clients. Similarly, a drop in demand at any given point can propagate in the opposite direction. This was the scenario that was proposed in Long and Plosser's seminal paper [169], spurring much of the research I describe below. For historical reasons, it should be noted that *input-output analysis* was pioneered by Leontief [165], inspired by ideas from Quesnay's *Tableau économique* [223].

Long and Plosser's framework was recently taken up by Acemoğlu *et al.* as a way of explaining large aggregate fluctuations and complementing the granularity scenario [4]. In their model, there are N firms or sectors, with the production of i given by a Cobb-Douglas production function, as

$$\pi_i = z_i^{\alpha} \ell_i^{\alpha} \prod_j Q_{ij}^{(1-\alpha)w_{ij}}, \tag{4.8}$$

where z_i is a *productivity shock*, which in their interpretation means a momentary increase/decrease in the effectiveness of the firm. For the other variables, ℓ represents labour, α is the share of labour in i's production, Q_{ij} is the amount of good j intervening in i's production and w_{ij} a weight coefficient defining the *input-output table*. The input-output table is also represented by a matrix $\mathbf{W}$ such that $(\mathbf{W})_{ij} = w_{ij}$, and it is assumed that $\sum_j w_{ij} = 1$. For a more detailed explanation of Cobb-Douglas and the more general family of CES production functions, see Appendix D. A crucial aspect from this production function, as stressed in the Appendix, is that goods can be substituted: if there is a drop in an input as $Q_{ij} \to f Q_{ij}$, with $f < 1$, then the production level π_i need not drop if another input Q_{ik} is multiplied by $f^{-\frac{w_{ik}}{w_{ij}}}$.

The labour in this model is supplied by a representative household for a wage w. It can consume a quantity C_i of good i, that is set by maximising an utility function of the form

$$\mathcal{U} = A \sum_i \log(C_i), \tag{4.9}$$

where A is a normalization constant that sets the output level of the economy.[6]

6 Note that in the original article the authors chose an utility $\mathcal{U} = A \prod_i (C_i)^{1/N}$, but this does not change the discussion below.

With this production function, it is straightforward to set *competitive equilibrium* conditions, where firms maximise their profits, households maximise their utility and markets clear, meaning that all the goods that are produced by a firm are bought by its clients, be it other firms or households. These conditions then give equations for the prices p_i of each good and for the output levels π_i. Crucially, under the Cobb-Douglas hypothesis these equations *always* have well-defined solutions, and an equilibrium always exists. This is now always the case, as the reader will see in Chapter 5.

Working everything out, the authors show that the GDP S of the economy is given by

$$\ln(S) = \sum_i v_i \varepsilon_i, \tag{4.10}$$

with $\varepsilon_i = \ln(z_i)$ and v_i the *influence vector* that measures the equilibrium share of sales of firm i, namely

$$v_i = \frac{p_i \pi_i}{\sum_j p_j \pi_j} = \frac{\alpha}{N} \sum_j (\mathbf{L})_{ij}, \tag{4.11}$$

where $\mathbf{L} = (\mathbf{1} - (1-\alpha)\mathbf{W})^{-1}$ is called the *Leontief inverse*.

There is then a clear parallel between Eqs. (4.4) and (4.10) if one compares the logarithm of the productivity shock ε_i with $\ln(1 + m + \sigma\eta_i)$. Ref. [4] then sees the input-output table $\mathbf{W}$ as the (weighted) connectivity matrix of a network, and they then prove that if the network is very unbalanced, with e. g. a power-law degree distribution or a "star" economy where one firm supplies everyone else, then the distribution of v_i is fat-tailed.

As a very simple toy example, one can consider a star network as depicted in Figure 4.4. For this network the matrices $\mathbf{W}$ and $\mathbf{L}$ read

$$\mathbf{W} = \begin{pmatrix} 0 & 1 & 1 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \mathbf{L} = \frac{1}{N} \begin{pmatrix} \alpha & 1-\alpha & \dots & 1-\alpha \\ 0 & \alpha & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha/N \end{pmatrix}, \tag{4.12}$$

and so the shares are $v_i = \frac{\alpha}{N} + (1-\alpha)\delta_{i,1}$, with a clear interpretation: every firm except for firm 1 sells its good only to the households, that consume a quantity $C_i \propto 1/N$. On the other hand, firm 1 must produce a fraction $C_1 \propto 1/N$ for households but must also supply a quantity $Q_{j1} \propto C_j \propto 1/N$ to every firm j, and so it ends up producing a quantity of order 1 to sell to other firms.

Taking the variance of Eq. (4.10) and supposing that the shocks ε_i are independent and of variance σ^2 leads to roughly the same situation as in Eq. (4.5), with $H = \sum_i v_i^2$ the herfindahl index of the economy. More complicated network structures are considered by the authors of Ref. [4], but they always need an unbalanced structure, with e. g. power-law distributed degrees, to have similar statistics of the shares v_i and therefore a herfindahl index close to 1, resulting in aggregate fluctuations that scale down much slower than $1/\sqrt{N}$.

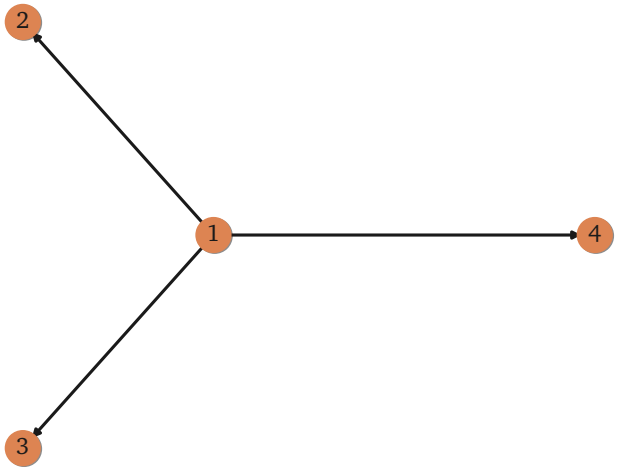


Figure 4.4: An example of a very unbalanced network: the star economy. Here, firm 1 is the sole supplier of all other firms, and it can be shown that within the Cobb-Douglas framework its share of sales v_1 is of order 1, while the other firms have sales of order $1/N$.

5

MARGINALLY STABLE PRODUCTION NETWORKS

In the previous section I have shown two explanations for the *small shocks, large business cycle* puzzle, as it was called by Ben Bernanke [44]. The first one, relying on "granularity" and explained in Sec. 4.1, relies too heavily on an assumption of independent shocks, a hypothesis that isn't empirically verified [114, 2]. The second one, describing a network of firms/sectors with Cobb-Douglas technologies, requires a very unbalanced production network to in fact *explain granularity* with a network origin, and so their mechanism for large aggregate fluctuations ultimately rests on granularity too.

A different clue is to consider instead that many shocks affecting the economy are actually of *endogenous* nature, and that the economy may self-organize into an unstable and turbulent state. This idea was mentioned in 1948 by Hawkins [142], and picked up again by Bak, Chen, Scheinkman and Woodford [34, 231], relying on the Self-Organized Criticality (SOC) ideas discussed in Section 3.4.4. As the proverbial sandpile that keeps growing into an unstable position prone to large avalanches, the economy, they argue, tends to flow into states in which it is anomalously susceptible to amplifying small shocks.

Similar ideas emerged also in theoretical ecology, with Robert May's seminal paper *Will a large complex system be stable?* [179] and recent follow-ups using ideas and methods coming from the statistical mechanics of disordered systems [47, 75, 74, 228]. For example, a very recent take on these ideas by Fyodorov and Khoruzhenko [117] considers very general evolution equations of the form

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = f_i\left(x_1, \ldots, x_N\right), \tag{5.1}$$

where $\vec{x}$ are the variables determining a system, and the vector field $\vec{f}(\vec{x})$ determines its (possibly random) variations.

Such equations can be studied on their own with relatively mild assumptions on the field $\vec{f}$, and linearising about a fixed point satisfying $\vec{f}(\vec{x}^\star) = \vec{0}$ this evolution can be written as

$$\frac{\mathrm{d}\vec{y}}{\mathrm{d}t} \approx \mathbf{A}\vec{y}, \quad \vec{y} = \vec{x} - \vec{x}^\star, \tag{5.2}$$

with the matrix $\mathbf{A}$ the Jacobian matrix at the fixed point,

$$(\mathbf{A})_{ij} := A_{ij} = \left[\frac{\partial f_i}{\partial x_j}\right]_{\vec{x}=\vec{x}^\star}. \tag{5.3}$$

Going back to May's simplified version, he suggested, along the lines of Ref. [126], to consider the variables x_i as the populations of N species in

an ecosystem, and taking the diagonal $A_{ii} = -1$. This amounts to setting a time-scale and considering that all species are roughly equivalent. The only important assumption here is that these terms are negative, and that therefore the point $\vec{x}^\star$ is dynamically stable in absence of inter-species interactions $A_{ij}, \quad i \neq j$.

Under these hypotheses, the matrix $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{J} - \mathbf{I}$, with $\mathbf{J}$ a matrix with a 0-diagonal and quantifying all of the intra-species interactions and $\mathbf{I}$ the identity matrix. The system is dynamically stable if for every eigenvalue λ of $\mathbf{A}$ we have $\mathrm{Re}\,\lambda < 0$. The next step is to *assume nothing* [71] on the interactions and take $\mathbf{J}$ as a random matrix. This follows the idea initiated by Wigner [273, 272] of modelling the interactions of complex systems (in his case, the hamiltonians describing large nuclei) as random matrices.

Appendix C gives an intuitive review on the basics of random matrix theory, but the heart of the argument is that the spectral radius of $\mathbf{J}$ grows with the number of species in the system, the variability of their interactions or the number of species with which each species interacts. The growing complexity of the system therefore places it closer and closer to an unstable regime.

The following model presents similar ideas, as an original contribution I published with Jean-Philippe Bouchaud in [191] and extending the network economy models presented in Sec. 4.2. The key difference is the introduction of more general production functions than for the Cobb-Douglas case, with the main characteristic that production inputs cannot be easily substituted. As a consequence, the existence of a competitive equilibrium, such as the one discussed at length in Ref. [4], is not always guaranteed, but depends on a constraint on the spectrum of the matrices describing the interaction between firms, called the *Hawkins-Simon* condition [142, 143, 112, 217, 220].

5.1 DESCRIPTION OF THE MODEL

The model considers N firms with a given input-output network determined by the technology available to firms. This "technology network" is a directed graph where nodes represent firms and where a directed edge $j \to i$ exists if i needs goods produced by j for its own production. Note that this framework allows for self-loops, with an edge $i \to i$ existing if a firm produces one of its own inputs. The node $i = 0$ conventionally represents households, and supplies firms with labour while consuming a part of their output. The link $j \to i$ carries a "stoichiometric weight" J_{ij}, measuring the number of j goods needed to make an unit of i's production. The set of suppliers of i is thus given by $\{j / J_{ij} \neq 0\}$, while the set of clients is $\{j / J_{ji} \neq 0\}$.

The production of i, π_i, is given by a so-called CES (Constant Elasticity of Substitution) function, which reads (see Ref. [224] and Appendix D)

$$\pi_i = z_i \left(\sum_j a_{ij} \left(\frac{J_{ij}}{Q_{ij}} \right)^{\frac{1}{q}} \right)^{-q}, \quad \text{with} \quad \sum_j a_{ij} = 1, \tag{5.4}$$

where z_i is the firm's productivity, Q_{ij} is the number of goods firm i buys from firm j and $a_{ij} \geq 0$ are weight parameters. In comparison with Eq. (4.8), the parameter z_i represents the efficiency with which firm i converts a given set of inputs into its output, and not random "productivity shocks" affecting the firm.

The parameter q measures the global substitutability of the different inputs. When $q \to 0$, no substitutes are available and Eq. (5.4) reduces to the classical Leontief production function:

$$\pi_i = z_i \min_j \left(\frac{Q_{ij}}{J_{ij}} \right). \tag{5.5}$$

The Cobb-Douglas framework $\pi_i = z_i \prod_j (Q_{ij}/J_{ij})^{w_{ij}}$ exposed in Sec. 4.2 corresponds instead to $q \to \infty$ and is often used to describe the average aggregate production of economic sectors [169], or of the economy as a whole. As stressed in Sec. 4.2, in a Cobb-Douglas economy the loss of a fraction f of good j can always be compensated by an increase of any other good k by a factor $1/f^{a_{ij}/a_{ik}}$. In the Leontief case, on the other hand, the loss of a fraction f of good j cannot be compensated and translates to an immediate loss of the same fraction f of total production π_i.

The Leontief case therefore models a situation where redundancy is costly. Firms therefore choose their suppliers with parsimony and cannot "rewire" (meaning find alternative suppliers) on short time scales in the real economy. For example, in the aftermath of the 2011 tsunami and Fukushima Daiichi nuclear power plant disaster, the shortage of a few seemingly unimportant components had a severe impact on the worldwide car industry [225, 256], very dependent on products manufactured in Japan.

The incident highlighted how competition led firms to have a very tight supply-chain strategy, as in the words of an observer: "In the race to provide better quality at lower prices, manufacturers picked very narrow, optimized supply chains" which caused them to be very dependent on the "one supplier that had the best product at the lowest price" [113]. At the time of the disruption, firms had to swiftly re-think their supply-chain strategy and large-scale rewirings of the production network took place. The influence of possible rewirings is beyond the scope of this paper, but for preliminary work in that direction see [90].

Furthermore, the Covid-19 economic and sanitary crisis is unfolding as I am writing this manuscript. The combination of lock-downs, border closings and uncertainty about the future has created a "combination of disaggregated supply and demand shocks" that propagate through the

supply chain [35], with many sectors running tight on their suppliers (see also [175]). It is not yet clear how recovery will take place, and what policy should be adopted to try and "jump-start" the economy [213]. Much debate has also taken place to criticise the supply-chain optimisation that in many cases left countries without essential goods like medical supplies [204], with many advocating in favour of resilience over pure optimality.

In the following, I will for simplicity focus on the extreme case of a Leontief production function $q \to 0$, but will show that the results hold true for in a range $q \in [0, q_c]$, where the critical value q_c depends on the network and on the productivities.

Calling p_i the price of the goods produced by firm i, its profit $\mathscr{P}_i$ reads:

$$\mathscr{P}_i = p_i \pi_i - \sum_{j \neq 0} Q_{ij} p_j - Q_{i0} p_0, \tag{5.6}$$

where p_0 is the labour wage. Optimizing the profit with respect to all inputs Q_{ij} (including labour Q_{i0}, which was denoted ℓ_i in Sec. 4.2) leads, for $q \to 0$, to the condition:

$$\forall (i,j), \exists \gamma_i \geq 0 \quad \text{s.t.} \quad Q_{ij} = \gamma_i J_{ij} \tag{5.7}$$

which can also be interpreted as saying that given an output level $\gamma_i := \pi_i / z_i$, the optimal choice for inputs Q_{ij} is to pick them proportionally to their stoichiometric weight, as buying more would result in waste. In this case, profit can be written as

$$\mathscr{P}_i = \gamma_i \left(z_i p_i - \sum_{j \neq 0} J_{ij} p_j - J_{i0} p_0 \right). \tag{5.8}$$

I also take the households' optimal consumption of good i, given a certain utility function and a vector of prices, to be given by $C_i > 0$[1].

As standard in the literature, and as done in Ref. [4], I now assume:

- Market Clearing, i. e. every good that is produced is either consumed by households or bought by other firms for their own production. Hence:

$$\pi_i = \sum_{j \neq 0} Q_{ji} + C_i \longrightarrow z_i \gamma_i - \sum_{j \neq 0} J_{ji} \gamma_j = C_i \quad (> 0). \tag{5.10}$$

- Competitive Equilibrium, i. e. competition drives profits to zero. Hence:

$$\mathscr{P}_i = 0 \longrightarrow z_i p_i - \sum_{j \neq 0} J_{ij} p_j = V_i \quad (> 0), \tag{5.11}$$

1 For example, for a utility function $\mathscr{U} = \sum_i \theta_i \log(C_i)$ and a certain budget B, the optimal consumption C_i is:

$$C_i = \frac{B}{\sum_j \theta_j} \frac{\theta_i}{p_i} := \frac{\Gamma_i}{p_i} \tag{5.9}$$

but any other type of utility function would work in this model, as long as consumption levels are strictly positive.

where $V_i := J_{i0}p_0$ after imposing $\gamma_i \neq 0$, $\forall i$ (otherwise Eq. (5.10) cannot be satisfied). One could also model firms attempting to impose mark-ups to reach a positive profit equal to a fraction φ_i of its sales $z_i\gamma_i p_i$. This simply amounts to shifting z_i to $z_i(1-\varphi_i)$ in Eq. (5.11).

Now, in order for the equilibrium to make sense, the solutions to Eqs. (5.10,5.11) must be such that $\gamma_i > 0$ and $p_i > 0$, $\forall i$; i. e. that equilibrium prices and quantities must be strictly positive. As first noted by Hawkins & Simon [143], this is not automatic and requires the matrix $\mathbf{M}$, defined by $(\mathbf{M})_{ij} = z_i\delta_{ij} - J_{ij}$ to be a so-called "M-matrix"[2], i. e. such that *all its eigenvalues have non-negative real parts* [112]. Therefore some conditions on productivities and linkages must be fulfilled for the economy to work.

5.1.1 *A link with theoretical ecology*

This condition is the equivalent, in an ecological context, of May's stability criterion that allows the equilibrium population of all species to be strictly positive [75, 74]. Rather interestingly, Eqs. (5.10,5.11) are identical, *mutatis mutandis*, to the equation determining the equilibrium size of species in a generalized Lotka-Volterra model [47]. Within that model, the evolution of the population N_i of species i is given by:

$$\frac{\mathrm{d}N_i}{\mathrm{d}t} = N_i\left(\mu_i - K_iN_i + \sum_{j\neq i}\alpha_{ij}N_j\right) \tag{5.12}$$

allowing one to introduce an interaction matrix $(\mathbf{A})_{ij} = K_i\delta_{ij} - \alpha_{ij}$, where δ_{ij} is the Kronecker delta, equal to 1 for $i = j$ and to 0 elsewhere. The terms μ_i are the species' different reproduction rates, diagonal terms of the matrix, K_i, quantify intra-species interaction and how the population of species i would saturate in absence of other species, and the α_{ij} off-diagonal terms quantify intra-species cooperation/competition terms.

In [47] the values of α_{ij} are taken to be i.i.d. gaussian variables of average μ/S and variance σ^2/S that are imposed to be symmetric, e. g. $\alpha_{ij} = \alpha_{ji}$, in order to facilitate analytical treatment. No further structure is imposed on the matrix $\mathbf{A}$, which is thus a full matrix.

Within that model, the stationary equilibria of Eq.(5.12) are given by:

$$\begin{cases} N_i = 0 \\ \quad \text{or} \\ \mu_i = \sum_j (\mathbf{A})_{ij} N_j \end{cases} \tag{5.13}$$

which shows the two possible paths of the dynamics: either all initial species are viable and survive to the stationary state or some of them must go extinct in order for an equilibrium state to emerge.

2 Note that if $\mathbf{M}$ is an M-matrix, $\mathbf{M}^t$ is also an M-matrix. An interesting property of an M-matrix is that all the elements of its inverse are non negative.

The dynamic stability of the model is also linked to the $N \times N$ matrix $\mathbf{A}$, or rather to the $N^* \times N^*$ sub-matrix $\mathbf{A}^*$ where one has removed the lines and columns that correspond to species that have gone extinct, and only retained the N^* remaining species. Mapping their model into a spin-glass problem, the authors find that dynamic stability is achieved when $\mathrm{Sp}(\mathbf{A}^*) \subset \mathbb{R}^+$, and that if the initial matrix $\mathbf{A}$ has negative eigenvalues, the Lotka-Volterra dynamics eliminate species as to obtain a matrix $\mathbf{A}^*$ that *saturates* May's bound, meaning that its spectrum is positive but with an eigenvalue very close to 0. The ecosystem is then found to be marginally stable, and to be extremely susceptible to small perturbations.

In this last situation, the system is found to be in a regime with *full replica symmetry breaking* (fRSB), a "glassy" regime where there are exponentially many different equilibria, corresponding to different possible choices of the N^* species that make up the matrix $\mathbf{A}^*$. These are therefore non-trivial dynamics that have not yet taken hold in the economics literature. Note that an explicit attempt to introduce ideas of glassy physics into economics has been made in Ref. [237].

Going back to the network economy, taking a more general CES production function with $q \geq 0$, the competitive equilibrium equation reads:

$$(z_i p_i)^\zeta - \sum_{j \neq 0} a_{ij}^{q\zeta} \left(J_{ij} p_j\right)^\zeta = V_i \quad (>0), \qquad \zeta := \frac{1}{1+q} \tag{5.14}$$

which boils down to Eq. (5.11) when $q = 0$ (for a detailed proof, see Eq. (D.8) in Appendix D.3). Interestingly, setting $\widehat{p}_i = p_i^\zeta$, one finds again that the condition for an admissible equilibrium is that the matrix $(\widehat{\mathbf{M}})_{ij} = z_i^\zeta \delta_{ij} - a_{ij}^{q\zeta} J_{ij}^\zeta$ is an M-matrix.

Note that since $\sum_{j \neq 0} a_{ij} < 1$, the Perron-Frobenius theorem ensures that Cobb-Douglas networked economies (such as those considered in Ačemoglu, Carvalho et al. [4] and corresponding to $q \to \infty$), *always* have an admissible equilibrium, for any network and any productivities. Therefore, the type of shock propagation that takes place in this model has no counterpart in a Cobb-Douglas economy. This constitutes a major departure from the shock propagation mechanism in Sec. 4.2.

5.2 STABILITY CONDITIONS FOR MODEL NETWORKS

Here and below I will for simplicity focus on the Leontief case, commenting on the more general case $q > 0$ in the conclusion. In order to gain some intuition on the stability conditions, consider first a random directed network, where each supply link J_{ij} is equal to J with probability r and 0 with probability $1-r$, and where all N firms have the same productivity z.

The spectrum of $\mathbf{M}$ in this case is well known when $N \gg 1$ and $r \sim \mathcal{O}(1)$. It consists of an isolated eigenvalue $\lambda_{\min} = z - JrN$ and a "sea" of complex eigenvalues uniformly distributed in a disc of radius $J\sqrt{r(1-r)N}$ centred

at z (see e. g. [132, 260], and compare with the symmetric case discussed in Sec. C.1.3).

The stability condition therefore reads $z > JrN$, i. e. productivity must be large enough for the economy to function. The most unstable eigenvector, corresponding to eigenvalue $\lambda_{\min}$, is the uniform vector $(1/\sqrt{N}, \dots, 1/\sqrt{N})$. As will be clear below, this corresponds to a case where crises are system wide. The same qualitative result holds when productivities are weakly heterogeneous, i. e. $z_i = z(1+\varepsilon_i)$ with $\varepsilon_i \ll 1$ (albeit $\lambda_{\min}$ is slightly shifted downwards by an amount $O(z^2\varepsilon^2/JN)$).

More interesting — but more complex! — is the case where the average number of suppliers $c = rN$ is of order unity (i. e. when $r = \mathcal{O}(N^{-1})$). In the *random regular network* (RRN) where each firm has exactly c suppliers (and c customers) chosen randomly among the $N-1$ other firms, one knows that the spectrum of $\mathbf{M}$ again consists of an isolated eigenvalue $\lambda_{\min} = z - Jc$ and a "sea" of complex eigenvalues distributed in a disc of radius $J\sqrt{c}$ centred at z.[3] For further intuition in the case where the matrix is symmetric, i. e. $J_{ij} = J_{ji}$, and when all firms have the same number of clients/suppliers see Appendix C.

When heterogeneity is introduced, either topological (i. e. letting the number of suppliers/customers to vary) or because the couplings J and the productivities z fluctuate, there are no exact results available, in particular in the case where $\mathbf{M}$ is not a symmetric matrix — see [183] for a very recent survey.

When $\mathbf{M}$ is symmetric, exact results are still scarce but a huge amount of work has been done in the physics, mathematics and computer science literature to characterize the eigenvalues and eigenvectors of such random matrices [272, 46, 108, 227, 118, 161, 43, 198, 48, 5]. The reason is that such symmetric random matrices appear is many physical situations, such as the vibration spectrum of amorphous solids or the energy spectrum of quantum systems with impurities. Such random matrices also appear in graph theory and computer science. The problem of estimating the extremal eigenvalue is of special importance, as it appears in many different problems (such as epidemic or rumour spreading [81], see also [253] — or crisis propagation as in the present work); the associated eigenvector is related to the concept of node centrality in network theory [199]; see also [80] and refs. therein.

In the sequel, Δ^2 denotes the variance of fluctuations (of connectivity, productivity, etc.). From the host of results accumulated in the last decades, the following general scenario is expected (see [48] and Fig. 5.1 for the case of random regular graphs):

- For $\Delta = 0$, all eigenvalues except one have their real part confined in a certain interval $\mathscr{I}$ ($= [z - \sqrt{c}, z + \sqrt{c}]$ in the RRN example), while the isolated eigenvalue is located to the left of this interval, at a non-zero distance g from its edge.

3 In this case, however, the density of complex eigenvalues is not uniform but is given by $\rho(\lambda) \propto (c^2 - |\lambda|^2)^{-2}$ for $|\lambda| < \sqrt{c}$ [183].

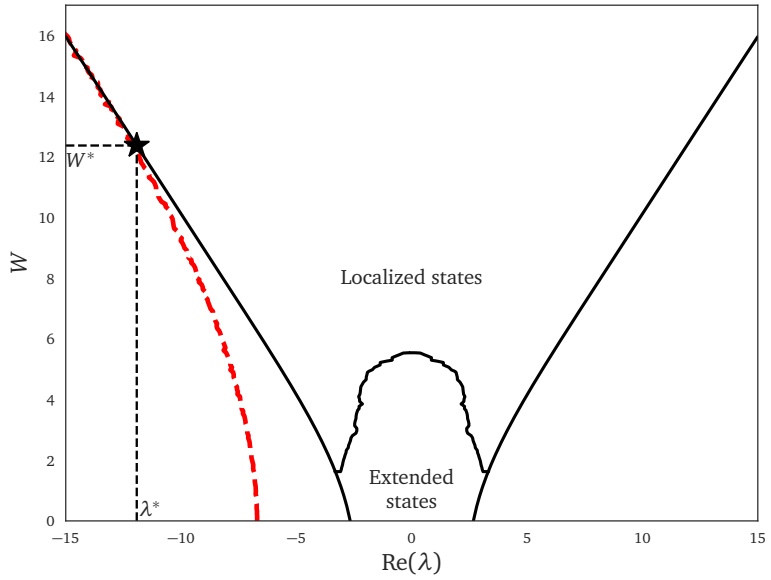


Figure 5.1: Numerical results for the structure of the eigenstates of a directed random regular network (RRN) with $N = 2000$ firms with $c = 7$ suppliers and clients each and $J = 1$. Productivities z_i are uniformly distributed in an interval $[z - W/2, z + W/2]$, but the x-axis is centred around z. Notice that the eigenstates contained in the bulk get localized as W increases. The left-most dashed red curve corresponds to the isolated eigenvalue that gets absorbed at the point marked by a black star in the graph, corresponding to values $W^* \approx 12.4$ and $\lambda^* \approx -12$. The boundary between extended and localized states is defined here by $H = 5/N$. Compare to Fig. 1 in [48], for the case of un-directed RRNs.

- As Δ increases, the interval $\mathscr{I}$ broadens and its edges become somewhat blurred, while the isolated eigenvalue gets closer and closer to the lower edge of $\mathscr{I}$ (see Fig. 5.1).

- Beyond a certain critical value Δ_c, the isolated eigenvalue is "eaten up" by $\mathscr{I}$ and disappears (this is called, in a different context, the Baik-Ben Arous-Péché (BBP) transition [30]). The mechanism through which this happens can be intuitively understood through the Dyson brownian motion, see Sec. C.2 and especially Figure C.5 for a very simplistic illustration of this.

Furthermore, as soon as Δ is non zero, the interval $\mathscr{I}$ is further subdivided into 3 intervals $\mathscr{I}_-, \mathscr{I}_0, \mathscr{I}_+$ (with $\mathscr{I}_0$ possibly empty, see Fig. 5.1), where the structure of the corresponding eigenvectors is markedly different. In the central part $\mathscr{I}_0$, eigenvectors are *extended*, or *delocalized*, whereas in the extreme parts $\mathscr{I}_-, \mathscr{I}_+$, eigenvectors are *localized*. In a hand-waving manner, "localized" means that most of the norm of the vector is concentrated on a few nodes (firms), whereas "delocalized" means that the norm

is well spread out over all nodes. More precisely, calling $v_1, v_2, \ldots, v_N$ the component of a normalized vector $|V\rangle$, the localized/delocalized nature of $|V\rangle$ is captured by its herfindahl index H (called Inverse Participation Ratio (IPR) in the physics literature, and related to that defined more broadly in Sec. 2.3.1):

$$H(|V\rangle) = \sum_i^N |v_i|^4. \tag{5.15}$$

As for the "normal" herfindahl index, a localized eigenvector is such that $H(|V\rangle) = \mathcal{O}(1)$ in the limit $N \to \infty$ whereas a delocalized eigenvector has $H(|V\rangle) = \mathcal{O}(N^{-1})$.

5.2.1 *A short digression on localization*

Indeed, the question of Anderson localization has been a major subject in condensed matter physics, as it gives a framework to study the conductivity properties of disordered materials. Traditionally, one considers a system of N atoms arranged in a lattice given by some graph $\mathcal{G}$, modelling metallic atoms in a crystal structure. The hamiltonian for a tight-binding model is usually written as follows

$$\mathcal{H} = -t \sum_{\langle i,j \rangle} (|i\rangle \langle j| + |j\rangle \langle i|) + \varepsilon \sum_i |i\rangle \langle i|, \tag{5.16}$$

and describes how electrons diffuse in the lattice. The first term corresponds to hopping between metallic ions (here $\sum_{\langle i,j \rangle}$ means that we are summing over neighbouring sites in $\mathcal{G}$) with a kinetic term $\propto t$, while the second term corresponds to the potential energy of an electron localized around ion i.

This hamiltonian can be represented by a real symmetric matrix $\mathbf{H}$ with components $H_{ii} = \varepsilon$ on the diagonal, corresponding to the energy of an electron in a given atomic site, and components $H_{ij} = -t$ if $i \leftrightarrow j \in \mathcal{G}$ and 0 elsewhere. The properties of the material under study are then reduced to the study of the eigenvectors and eigenvalues of these matrices, with the former being linked to the density of electrons in a given site and the latter being the energy of said electrons. For a complete introduction, see e. g. [9].

Given an eigenvector $|\psi\rangle$ of $\mathbf{H}$ and a site j represented by the vector $|j\rangle$, the quantity $|\langle j|\psi\rangle|^2$ is interpreted as the electronic density in that site. Most often in crystals the eigenvectors of the tight binding hamiltonian are given by standing waves $|\psi\rangle \propto \sum_{j \in \mathcal{G}} e^{ikj} |j\rangle$ where k is a constant linked to the corresponding eigenvalue. In that case, the quantity $|\langle j|\psi\rangle|^2$ is non-zero for all sites, and the material can conduct electricity.

That being said, real materials have some level of disorder in them, either through geometric defects where the atoms' actual positions are shifted with respect to the crystal structure, or through impurities, i. e. atoms of another kind that are scattered around and introduce randomness in the

hamiltonian. This can be done by changing $\varepsilon \to \varepsilon_i$, and setting the energy level of each ionic site as a random quantity. In 1958 P.W. Anderson [16] showed that in some cases the introduction of randomness into the system prevents the electrons from successfully diffusing in the lattice, and leads to so-called Anderson localization.

In particular, Biroli, Semerjian and Tarzia [48] studied Anderson localization in Bethe lattices, which are in the large N limit equivalent to random regular graphs. Their model consists of determining the eigenvectors and eigenvalues of a matrix $\mathbf{H} = (H_{ij})$ such that $H_{ij} = -t$ if the link $i \leftrightarrow j$ exists in the graph and $H_{ii} = \varepsilon_i \in [-W/2; W/2]$, and can be thought of as the study of the eigenstates of a symmetric Leontieff matrix, with $J_{ij} = J_{ji} = J$ and where each firm has a random productivity taken from an uniform distribution in an interval of width W and is connected to exactly $k+1$ other firms. Figure 5.1 is in fact inspired from this work, but as a non-symmetric counterpart.

For any non-zero value of W both edges of the bulk of the spectrum correspond to localized states, with a finite value of their corresponding IPR in the large N limit. The isolated eigenvalue persists for small values of W, but after a certain critical value it is absorbed into the bulk, which still has an edge consisting of localized states. This mechanism is closely related to the Baik-Ben Arous-Péché (BBP) transition [30] for hermitian matrices and can be understood using Dyson brownian motion [7, 8] as the hybridization of the eigenvector corresponding to the isolated eigenvalue with vectors in the localized edge of the spectrum.

Additionally, previous work by Biroli and Monasson [46, 187] have shown that, in the context of random tree-like and small-world networks, localized states appear at the edge of the spectrum, with the states in the middle being de-localized. Furthermore, localized states are induced by geometrical defects and correspond to sites with an abnormally high or low connectivity, as well as sites with a significantly different value of the diagonal component H_{ii}. Note also that there are recent results showing that the adjacency matrices of Erdös-Renyi graphs with a connectivity of order $\ln(N)$ have eigenvectors that split into a delocalized phase close to the middle of the spectrum and a semi-localized (i. e. in a few vertices) phase near the edges [10].

Although the results presented so far concern the localization properties of real symmetric matrices, there has been a considerable amount of work done recently on the localization properties of non-hermitian systems [197, 13, 258, 141]. Theoretical results have been obtained only for simple topologies (such as a ring graph, equivalent to the 1-d chain), but it is possible to transpose ideas from hermitian localization theory and use notions such as the IPR or herfindahl to quantify how localized the eigenvectors may be.

5.2.2 *Crisis propagation*

The importance of this distinction between localized and delocalized states for crisis propagation in the context of the model will become clear below.

Owing to the structure of **M**, the Perron-Frobenius theorem ensures that its leftmost eigenvalue $\lambda_{\min}$ is real with a real positive eigenvector $u_i > 0$. As stated above, $\lambda_{\min}$ must be positive for **M** to be an M-matrix, i. e. for all prices and all quantities to be positive. As $\lambda_{\min} \to 0$, the economy becomes more and more fragile to external shocks.

Let us for example consider the case where the productivity z_i of some firms decrease by $-\varepsilon\Delta_i < 0$, and/or that some of the stoichiometric weights J_{ij} increase by some amount $\varepsilon\Delta_{ij} > 0$. Using standard perturbation theory to first order in ε[4], one finds that the leftmost eigenvalue is shifted as:

$$\lambda_{\min} \longrightarrow \lambda_{\min} - \varepsilon \left[\sum_i \Delta_i u_i^2 + \sum_{i \neq j} \Delta_{ij} u_i u_j \right]. \tag{5.17}$$

Since both correction terms are negative, this formula shows that as the system becomes marginally stable, any local decrease of productivity/increase of required inputs tips the system towards the unstable region. A certain number of prices/quantities then become negative. Intuitively, the physiognomy of these "crises" will depend on the localized/delocalized nature of the eigenvector $|U\rangle$ corresponding to $\lambda_{\min}$, as discussed next.

The next order correction to Eq. (5.17) is of order ε^2/g, where g is the gap between $\lambda_{\min}$ and the next eigenvalue of **M**; therefore first order perturbation theory is only valid provided $\varepsilon \ll g$. Now, two cases must be distinguished, depending on the strength Δ of the heterogeneities:

- When $\Delta < \Delta_c$, the leftmost eigenvalue is isolated, in which case $g > 0$ even when $N \to \infty$. The first order result Eq. (5.17) is then valid when ε is small enough. Furthermore, the associated eigenvector $|U\rangle$ is delocalized. From Eq. (5.17), one deduces that a localized shock — say on firm ℓ alone — decreases $\lambda_{\min}$ by $\sim -\delta z_\ell / N$. The system is unstable when $\delta z_\ell > N\lambda_{\min}$, but for this condition to be compatible with $\delta z_\ell \ll g$, one must also require $N\lambda_{\min} \ll g$. When destabilized, the shock propagates over the whole system, because of the delocalized nature of $|U\rangle$. In the case of a small *global* productivity shock $\delta z_i = \delta z, \forall i$, the destabilisation occurs as soon as $\delta z > \lambda_{\min}$.

- When $\Delta > \Delta_c$, the leftmost eigenvalue is at the edge of the interval $\mathscr{I}_-$, such that the gap $g(N)$ generically goes to zero as $N \to \infty$. Furthermore, the associated eigenvector $|U\rangle$ is now localized, usually centred around particularly low productivity/high connectivity firms (called the Lifschitz regions in the physics literature [167, 262, 48]). In this case, however, first order perturbation theory breaks down as

4 Indeed, given an eigenvector $\vec{u}$ corresponding to an eigenvalue λ of a matrix **B** subject to a perturbation $\mathbf{B} \to \mathbf{B} + \varepsilon\mathbf{P}$, the first order correction to λ in epsilon reads $\lambda \to \lambda + \varepsilon^t \vec{u}\mathbf{P}\vec{u}$.

soon as $\varepsilon \sim g(N)$, so one must have recourse to numerical simulations to characterize the associated crisis patterns — see Sec 5.4 for a comparison with empirical data.

Finally, the following remarks should be useful to get an intuition about crisis propagation in the model. Note that one can write prices and outputs using the inverse matrix $\mathbf{M}^{-1}$ and its transpose. Hence, the price response to some generic perturbations δy (for example to productivity, household consumption, etc.) can be expressed using the eigenvalues and eigenvectors as:

$$\delta p_i = \sum_\alpha \ell_i^\alpha \frac{1}{\lambda_\alpha} \langle r^\alpha | \delta y \rangle , \tag{5.18}$$

where ℓ^α, r^α are, respectively, the left and right eigenvectors of $\mathbf{M}$ associated to eigenvalue λ_α. Similarly, for production

$$\delta \gamma_i = \sum_\alpha r_i^\alpha \frac{1}{\lambda_\alpha} \langle \ell^\alpha | \delta y \rangle . \tag{5.19}$$

In the limit where $\lambda_{\min}$ touches zero with a finite gap g, one can approximate these responses as

$$\delta p_i \approx \frac{\ell_i^{\min} \left\langle r^{\min} \middle| \delta y \right\rangle}{\lambda_{\min}}; \qquad \delta \gamma_i \approx \frac{r_i^{\min} \left\langle \ell^{\min} \middle| \delta y \right\rangle}{\lambda_{\min}}. \tag{5.20}$$

Hence, the amplitude of the response of prices depends on the overlap $\left\langle r^{\min} \middle| \delta y \right\rangle$ and is localized on the left eigenvector $\ell^{\min}$, and vice-versa for production. This will be illustrated using real data in Sec. 5.4.

In order to understand intuitively the divergence of the response to perturbations, consider the simple case where $\forall i, z_i = z$. One can expand $\mathbf{M}^{-1}$ in the stable region as

$$\mathbf{M}^{-1} = \frac{1}{z} \sum_{k=0}^{\infty} \left(\frac{\mathbf{J}}{z} \right)^k \tag{5.21}$$

with $(\mathbf{J})_{ij} = J_{ij}$, since the stability condition implies that the spectral radius of $\mathbf{J}$ is smaller than z. Now, the term $(\mathbf{J}^k)_{ij}$ consists of the sum of all paths of length k linking firm j to firm i. Marginal stability corresponds to this sum becoming divergent, with paths of all lengths contributing to $(\mathbf{M}^{-1})_{ij}$. This interpretation also holds in case of heterogeneous z_is.[5] The instability is therefore related to a situation where shocks can propagate over paths of arbitrary length in the input-output network. This is closely related to second-order phase transitions in physics, where correlations extend over macroscopic distances and the response to small perturbations diverges, see e. g. [234] and the discussion in Sec. 3.4.5.

5 Indeed one can always write an M-matrix as $\mathbf{M} = z_{\max}\mathbf{1} - \mathbf{B}$ where $\mathbf{B}$ is non negative, and expand the series as $\mathbf{M}^{-1} = \frac{1}{z_{\max}} \sum_{k=0}^{\infty} \left(\frac{\mathbf{B}}{z_{\max}} \right)^k$.

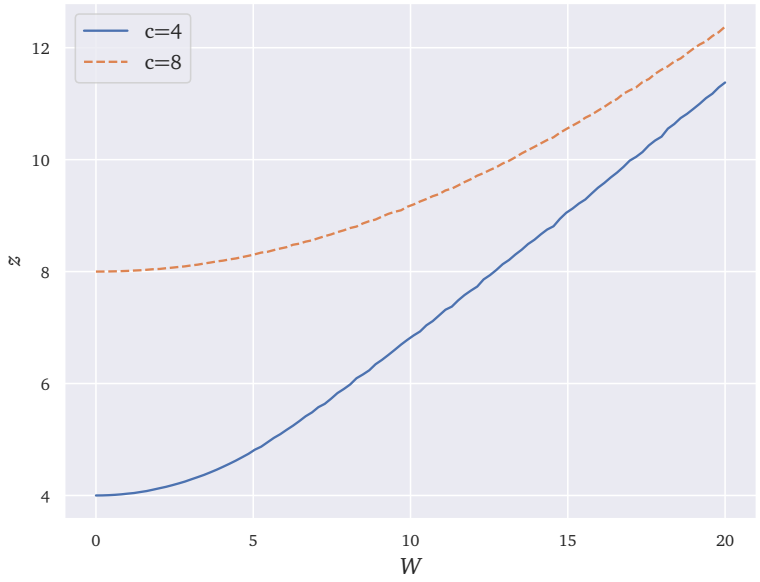


Figure 5.2: Plot of the average productivity $z = f_c(W)$ needed to stabilize the economy in the case of a directed RRN with productivities uniformly distributed in $[z - W/2; z + W/2]$, for connectivities $c = 4$ and $c = 8$. Notice that $f_c(0) = c$ and that there is a linear behaviour of $f_c(W)$ as $W \to \infty$, as expected. Economies with $N = 2048$ firms were simulated. Error bars are too small to be visible.

5.3 NUMERICAL RESULTS: BROAD DISTRIBUTION OF FIRM SIZES AND CRISES

This section considers the simplest model of a random regular network, where each firm has exactly c suppliers and c customers, each chosen randomly among the $N-1$ other firms (other types of networks will be discussed in [101]). Each firm has a random productivity uniformly chosen in the interval $[z - W/2, z + W/2]$, such that z is the average productivity and $\Delta = W/2\sqrt{3}$. The stoichiometric coefficients J_{ij} are taken to be all equal to J. Without loss of generality, J can be set to unity. In addition, I set $V_i = 1$ for simplicity and take the households' consumption $C_i = 1/p_i$ as obtained from a logarithmic utility function with identical preference for all products.

When $W = 0$, the spectrum of $\mathbf{M}$ can be computed, as discussed in the previous section, with an isolated leftmost eigenvalue given by $\lambda_{\min} = z - Jc$. As W increases, the spectrum evolves as shown in Fig. 5.1. In the case of $c = 7$, the isolated eigenvalue disappears when $W = W^* \approx 15.5$, and the edge of the spectrum corresponds to a localized eigenvector. In the following, I have set the average productivity to $z = f_c(W)$ such that $\lambda_{\min} = \varepsilon = 10^{-8}$ (the function $f_c(W)$ is shown in Fig. 5.2). The model then only depends on two parameters: the connectivity c and the

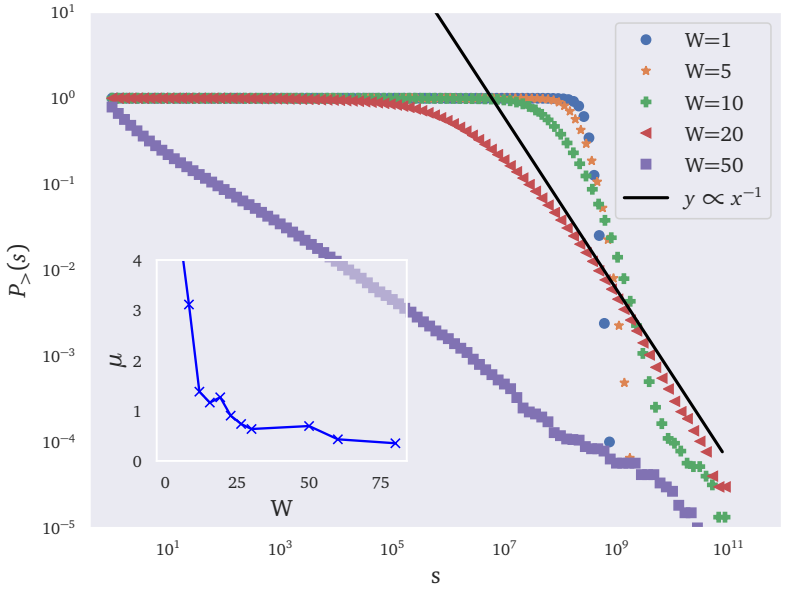


Figure 5.3: Log-log cumulative distribution of firm sizes $P_>(S)$ defined by sales $S_i = z_i \gamma_i p_i$, along with a curve corresponding to power-law $S^{-\mu}$ with exponent $\mu = 1$ (Zipf) for comparison. For small values of W the distribution of firm sizes is sharply peaked at a value of order $1/\varepsilon = 10^8$. Increasing W causes the distribution to get fatter tails with an apparent power-law exponent μ that decreases with W. Here, $c = 4$, $z = f_c(W)$, $N = 1500$. Inset: power-law exponent μ as a function of W, along the critical line of the model.

productivity heterogeneity W. In the following, I study the characteristics of the corresponding economy along the critical line.

Two quantities are of particular interest for this paper (a more throughout account of the results will be reported in [101]). One is the distribution of firm size, defined as the total sales $\mathscr{S}_i = z_i \gamma_i p_i$. Quite interestingly, while this distribution has thin tails when W is small, it becomes fat-tailed (Zipf-like) as W increases, as observed empirically [25], but with an exponent that appears to vary with W and c (see Fig. 5.3, inset). The emergence of a power-tailed firm size distribution is a consequence of the criticality of the model, but requires the extreme eigenvectors to be localized and heterogeneous, as it is the case for W sufficiently large.

The second quantity is the distribution of crises amplitudes $\mathscr{A}$, defined as the total size of the firms that are such that their equilibrium price becomes negative after an idiosyncratic shock of amplitude $-\delta z_\ell$ hitting a certain firm ℓ. (Shocks on the coefficients J_{ij} lead to qualitatively similar results). While in a fraction of cases nothing much happens, an avalanche can develop where a number of firms "go under", in the sense that their equilibrium price becomes negative. Conditioned to such events, the distribution of $\mathscr{A}$ is found to be of three types (see Fig 5.4):

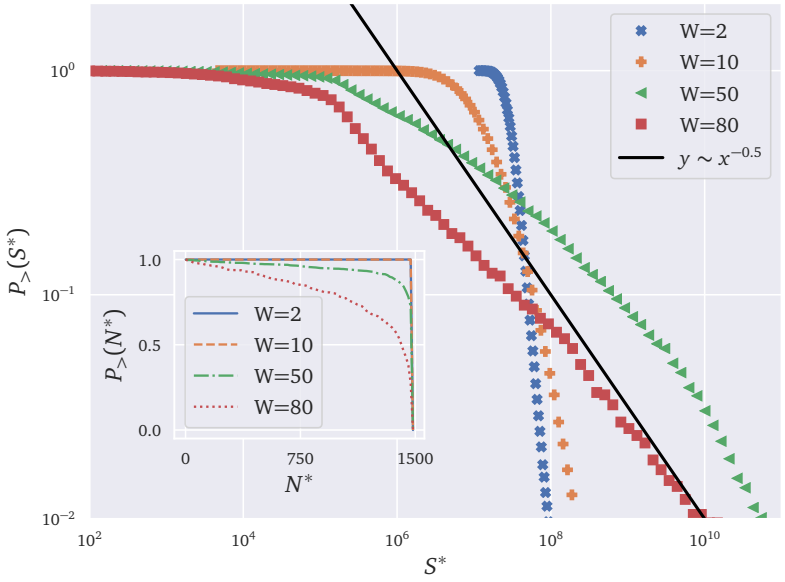


Figure 5.4: Log-log cumulative distribution of avalanche sizes as defined by the total sales of firms gone under after a shock, for $c = 4$, $z = f_c(W)$, $N = 1500$ and $\delta z_\ell = 0.05$. One again observes a broad, power-law tailed distribution of casualties for large enough W's. Inset: cumulative distribution of the number of firms N^* that have gone under after a shock. Notice that $W = 2$ and $W = 10$ correspond to mostly system-wide avalanches (i. e. $N^* \approx 1500$), while larger values of W correspond to avalanches of all sizes.

1. mostly "system wide", where a substantial fraction of the output is wiped out. This occurs when $W < W^*$ and $\delta z_\ell > N\lambda_{\min}$, as expected from the general discussion;

2. thin-tailed, where avalanches are restricted to particularly fragile firms connected to ℓ. This corresponds to $W > W^*$, and weak perturbations $\delta z_\ell < g$, in which case only one or a few localized eigenvectors close to the edge propagate the crisis.

3. fat-tailed, where small crises coexist with large crises (a feature of Self-Organized Criticality, as recalled in the introduction). This happens when $\delta z_\ell \gg g$, i. e. when a large collection of eigenstates are mobilized in the crisis propagation.

The generic existence of three crisis scenarii is quite interesting. In particular, the possibility that a small, idiosyncratic shock can lead to system-wide trouble, or else to avalanches of all sizes, has potentially deep consequences on our understanding of the business cycle and on crisis prevention policies. Of course, the above analysis postulates that the economy is close to criticality, i. e. that $\lambda_{\min} \to 0$. Why this should be the case is obviously the crux of the matter, and will be discussed in the next section.

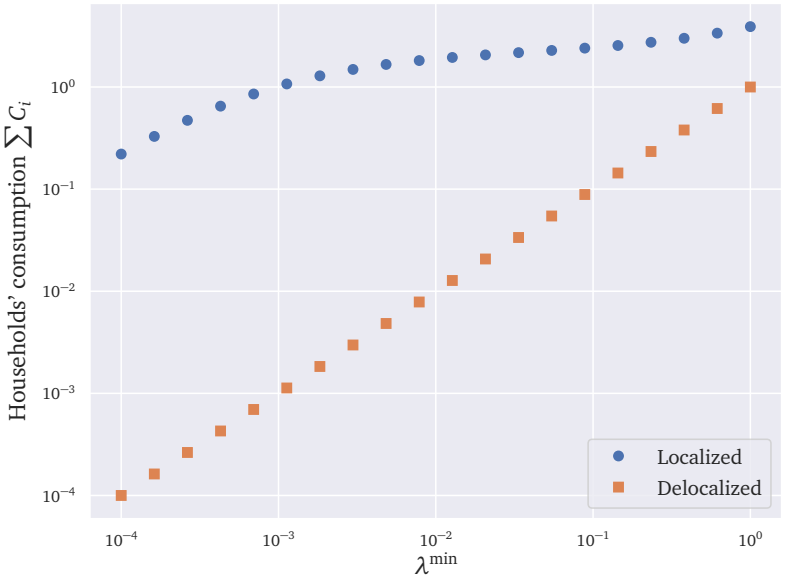


Figure 5.5: Total consumption of households vs. $\lambda_{\min}$ in the localized and delocalized cases. Intuitively, one expects the prices of goods represented in the eigenvector corresponding to $\lambda_{\min}$ to behave as $\lambda_{\min}^{-1}$, while a logarithmic utility function implies that consumption scales as the inverse of the price, leading naturally to the consumption of those goods to be proportional to $\lambda_{\min}$. In the delocalized case, all goods are concerned and thus global consumption plummets, as seen in the graph (orange squares). On the other hand, only the handful of goods associated to $\lambda_{\min}$ see their consumption decline in the delocalized case, corresponding to the blue dots. An economy with 1000 firms and a connectivity $c = 4$ was used for this plot, with $W = 0$ for the localized case and $W = 12$ for the localized case.

It is also interesting to plot the total consumption of households as a function of $\lambda_{\min}$ in the two cases above: delocalized ($W < W^*$) vs. localized ($W > W^*$) crises. In the first case, the whole economy grinds to a halt as $\lambda_{\min} \to 0$, as expected. In the second case, only a fraction of the total consumption (mostly coming from firms represented in the corresponding localized eigenvector) is affected. See Figure 5.5 for an illustration of this point.

This general scenario is strongly reminiscent of similar ideas in an ecological context, where the disappearance of a single species can lead to mass extinctions mediated by network effects [179, 47]. A major difference, however, is that the economic network is not static and can in principle adapt to new conditions on relatively short time scales (at least compared to evolutionary time scales). With this framework, one would expect that if a supplier undergoes some difficulty (i. e. its equilibrium production is found to be negative), its customers may choose to rewire and look for alternatives.

Furthermore, one expects that the market clearing and zero profit conditions will be temporarily violated. An extension of the present model that takes such dynamical effects into account would certainly be extremely interesting (see [52, 90] for preliminary work in that direction). But what is clear is that if rewiring takes time and/or is costly, the "paper crises" found above could indeed materialize as actual defaults, or at least acute difficulties. Since economic frictions are substantial and rewiring cannot be instantaneous, the present scenario could nonetheless be relevant to understand real world crises [225, 256, 113].

Let us finally come back to the case of partial substitutability, i. e. when the parameter q appearing in Eq. (5.4) is strictly larger than zero. Since the matrix $(\widehat{\mathbf{M}})_{ij} = z_i^{\zeta}\delta_{ij} - a_{ij}^{q\zeta}J_{ij}^{\zeta}$ is a continuous function of q, it is clear that if the smallest eigenvalue $\lambda_{\min}$ is negative for $q = 0$, it will remain so for a certain range of q. A numerical check on some examples shows that this is indeed the case; the economy is only stabilized when q exceeds a (problem dependent) value $q_c > 0$. Not surprisingly, allowing for more substitutability can stabilize an otherwise unfeasible economy.

5.4 REAL INPUT-OUTPUT NETWORKS

In this section I attempt to confront the previous model with available, but partial data. At this stage, this is more of an exercise that gives colour to the general framework.

Since the stability criterium depends on spectral properties of a matrix describing the entirety of the economy, a throughout analysis would in principle require highly detailed network data. Most available data, however, consists of input-output tables describing sale and purchase relationships between entities, be they firms or larger entities such as sectors or even countries, or simple relational data describing who is in a client-supplier relationship with whom, with the latter having a significantly larger coverage. While it allows to some degree to infer the importance of certain firms in the network, it does not correspond directly to the J_{ij} coefficients that appear in the formalism of the model.

The total productivity factors z_i, which measure the efficiency with which a firm turns inputs into outputs, are also hard to deduce from actual production data. In particular, detailed output data is seldom available, and so productivity measures must be constructed from revenue data only, allowing for potential errors between the actual and the inferred productivity levels. For a detailed discussion, see [255].

Detailed data is becoming increasingly available, and the purpose is now to show that spectral analysis of production networks is in principle possible.

5.4.1 *Dataset and definitions*

I've used the FactSet Supply Chain Relationships database to build a supply chain network. The FactSet dataset contains a list of relational data between firms, stating if firms A and B have a client/supplier relation, if they are in competition or if they have a joint venture. It is built by collecting information from primary public sources such as SEC 10-K annual filings, investor presentations and press releases, and covers about $23,000$ publicly traded companies with over $325,000$ relationships. Since the relationships are inferred from data released to the public, one cannot be sure that it is an exhaustive database of all the relationships between firms, but the subset of relationships deemed important by the firms themselves.

Such links between firms have a finite duration in time and have thus a beginning and end date. For this study, I have chosen the set of client/supplier relationships during the whole year of 2015. This allows to build a graph $\mathscr{G}$ where a link $i \to j$ exists whenever i is reported to be a supplier of j or when j is reported to be a client of i. This graph $\mathscr{G}$ consists of 237 weakly linked subgraphs [6], of which I pick the largest subgraph $\mathscr{G}_0$. The graph $\mathscr{G}_0$ consists of $10,447$ firms with $40,300$ relationships between them. The remaining subgraphs are very small and consist of at most a few tens of firms.

From $\mathscr{G}_0$ I then construct an adjacency matrix $\mathbf{A}$ defined by

$$(\mathbf{A})_{ij} = \begin{cases} 1 & \text{if } j \to i \in \mathscr{G}_0 \\ 0 & \text{otherwise} \end{cases} \tag{5.22}$$

which carries the topology one expects from the matrix $(\mathbf{M})_{ij} = z_i \delta_{ij} - J_{ij}$ defined in the main text. Indeed if for simplicity one supposes that all firms have the same productivity z and that $J_{ij} = 1$ for all links in the graph, then one has $\mathbf{M} = z\mathbf{1} - \mathbf{A}$. A full study of the properties of $\mathbf{M}$ would therefore require the supplementary knowledge of the values taken by the J_{ij} and z_i terms, i. e. to have access to the dollar amount of products exchanged between firms and the total production of each firm.

5.4.2 *Spectral study of the adjacency matrix*

As last sections suggest, it is interesting to look into the properties of the eigenvalues and eigenvectors of the matrix $\mathbf{M}$, which under the hypothesis presented above are the same as those of the matrix $-\mathbf{A}$, up to a shift by z on the real axis for the eigenvalues.

A notion introduced I have introduced is that of localized and delocalized eigenvectors. To quantify the localization properties I now introduce the

6 A weakly linked subgraph is such that any two nodes in it can be linked by a directed path.

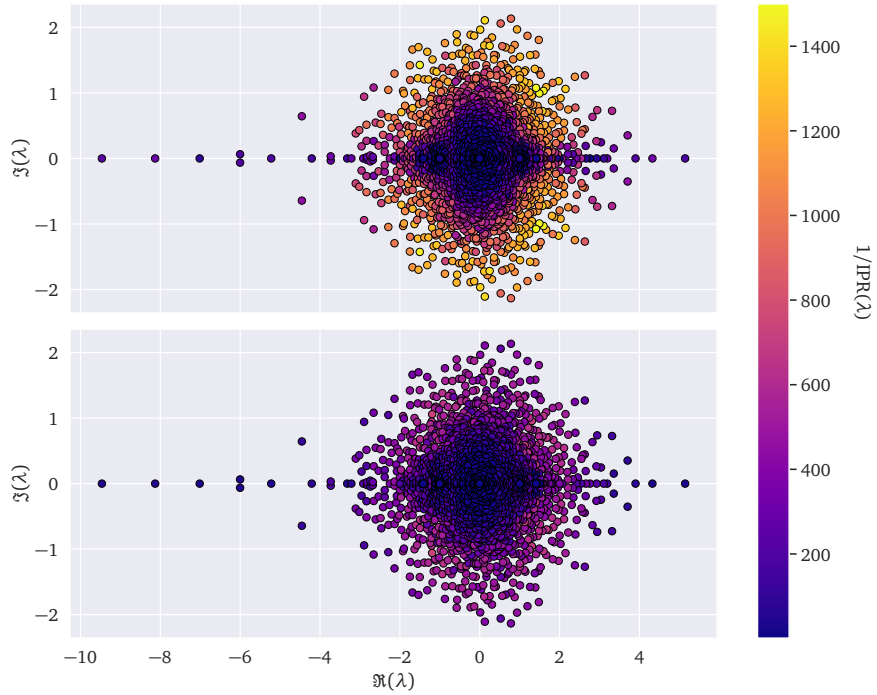


Figure 5.6: Spectrum of $-\mathbf{A}$ in the complex plane, with the colour of each eigenvalue λ given by $L(\lambda)$, defined as the inverse of the Inverse Participation Ratio H of the corresponding eigenvector, e. g. the number of firms over which the eigenvector is effectively spread. The colour on the top plot corresponds to the right eigenvectors, while the one on the bottom corresponds to the left eigenvectors. Notice the different localization profiles on left and right eigenvectors.

Inverse Participation Ratio, or IPR,[7] H of a normalized complex-valued eigenvector $\left|v^\lambda\right\rangle = \left(v_1^\lambda, \ldots, v_N^\lambda\right)$ associated to an eigenvalue λ as:

$$H(\lambda) := \sum_i \left|\left|v_i^\lambda\right|\right|^4. \tag{5.23}$$

Indeed, as discussed in Sec. 2.3.1, mind that for a perfectly delocalized eigenvector one should have $\forall i, \quad v_i^\lambda = 1/\sqrt{N}$ and therefore an $H(\lambda) = 1/N$, while an eigenvector localized on a single site $\left|v^\lambda\right\rangle = (1, 0, \ldots, 0)$ has trivially a value $H(\lambda) = 1$. It follows then that $L(\lambda) = 1/H(\lambda)$ is a measure of the number of sites over which a given eigenvector associated to the eigenvalue λ is spread out.

I have computed the eigenvalues and eigenvectors of $-\mathbf{A}$, as visible on Figure 5.6, where one can see qualitatively that the spectral properties of the real adjacency matrix are not far from the simplistic random regular graph case shown in Figure 5.1: there is indeed a bulk to the right hand side with isolated eigenvectors to the left of it, and with varying localization properties of the eigenvectors across the spectrum.

7 Akin to the Herfindahl index used in the economics literature.

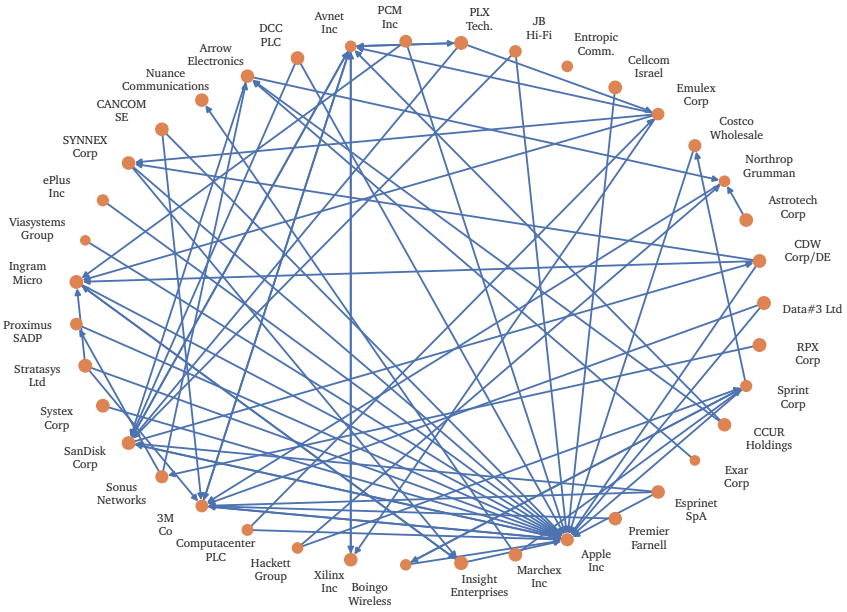


Figure 5.7: Subgraph of the 38 firms with the largest contributions to $\left|r^{\min}\right\rangle$.

Of capital interest to this study are the left-most eigenvectors $\left|l^{\min}\right\rangle$ and $\left|r^{\min}\right\rangle$, which owing to the Perron-Frobenius theorem have real positive components and are associated to a real eigenvalue $\lambda_{\min}$, as can also be seen in Figure 5.6. In the graph under analysis, I found that the eigenvectors are spread out over $L_r(\lambda_{\min}) \simeq 40$ firms for the right eigenvector and $L_l(\lambda_{\min}) \simeq 185$.

For both eigenvectors, I have listed the 20 firms with the most important contributions on Tables 5.1 and 5.2, as well as their yearly reported sales for the year 2015[8]. I have also represented the subgraph of the 38 firms with the largest contributions to these eigenvectors and their inter-linkages in Figures 5.8 and 5.7. This allows for a better understanding of Eq. (5.20): if a shock hits firms represented in Figure 5.7 this will be reflected in the prices of goods produced by firms represented in Figure 5.8.

This corresponds fairly well to basic intuition, as firms in Figure 5.7 correspond roughly to firms producing goods (such as electronics and software) that one expects to be purchased, after possible transformations along the supply chain, by retail and communication firms in Figure 5.8. The same holds by the inverting the roles of the firms, as idiosyncratic shocks to firms contributing to the left eigenvector will have an effect on the output of firms represented in the right eigenvector.

Although the data available to us is not as complete as one could wish for a full study of the matrix $\mathbf{M}$, I think nonetheless that it supports some of the qualitative conclusions of the model. For example, I've argued that

8 In a few cases this data was not available.

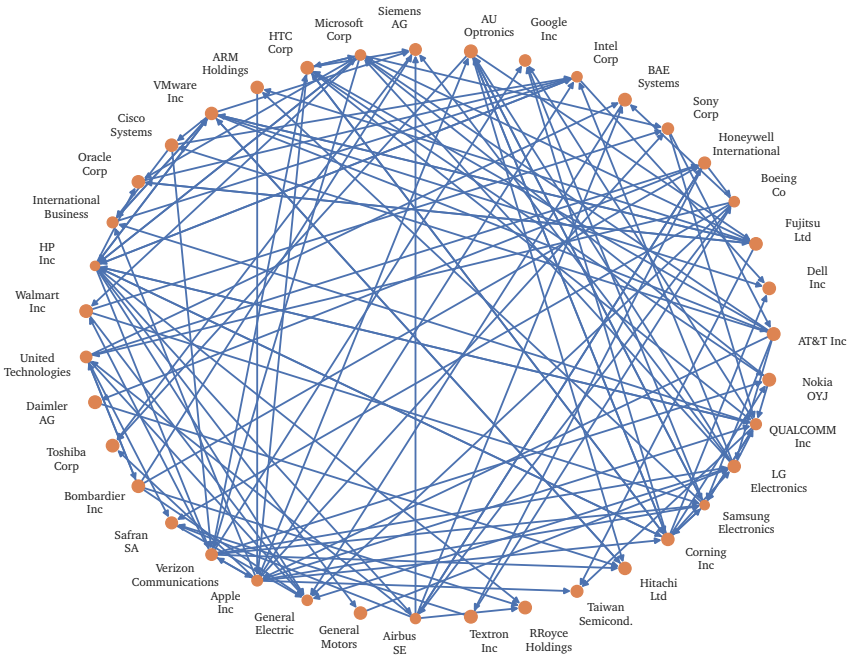


Figure 5.8: Subgraph of the 38 firms with the largest contributions to $\left|l^{\min}\right\rangle$.

the power-law firm size distribution could be a feature explained by the proximity of an instability, with the largest firms being those that have a high overlap with the leftmost eigenvectors of $\mathbf{M}$. To test this, I have plotted the sales of the 36 firms with the largest contribution to $\left|r^{\min}\right\rangle$ and 34 firms with the largest contribution to $\left|l^{\min}\right\rangle$ against their contributions of these firms to the IPR/Herfindahl of these vectors[9] in Figure 5.9, suggesting an increasing relation between the overlap of a firm with $\left|v^{\lambda_{\min}}\right\rangle$ and its sales.

5.5 MARGINAL STABILITY: DISCUSSION & CONCLUSION

In this section, I will motivate the claim that generic economies — like many other complex systems, see e. g. [32, 103, 84, 195, 47] — might "self-organize" to sit, at least temporarily, close the boundary of the stable region, i. e. satisfy the marginal stability criterion $\lambda_{\min} \to 0$. Several types of evolutionary forces act to that effect. One is simply the creation of new firms, that lead to an *effective* reduction of productivity and increase of connectivity. To see this, consider that the economy consists of N firms in equilibrium and add an additional firm indexed by $\star$, with productivity $z_\star$, labour requirements $J_{\star 0} = V_\star / p_0$ and links $J_{\star i}, J_{j\star}$ to the N pre-existing firms. The equilibrium condition for price $p_\star$ is:

$$p_\star = \frac{V_\star}{z_\star} + \sum_{j=1}^{N} \frac{J_{\star j}}{z_\star} p_j. \tag{5.24}$$

9 Which is none other than the quartic root of the scalar product between a firm's vector and $\left|l, r^{\lambda_{\min}}\right\rangle$.

Company Name	**Sales (kUSD)**	**Herfindahl Contribution** $\|\|r_i^{\lambda_{\min}}\|\|^4$
HP Inc	1.40e+06	3.62e-03
Samsung Electronics Co Ltd	1.37e+06	3.21e-03
Airbus SE	6.69e+05	1.10e-03
Boeing Co/The	1.12e+06	9.37e-04
General Electric Co	1.85e+06	9.28e-04
Intel Corp	4.99e+05	8.85e-04
Microsoft Corp	5.61e+05	8.74e-04
Apple Inc	2.20e+06	5.75e-04
Intl. Business Machines Corp	7.93e+05	4.53e-04
QUALCOMM Inc	1.57e+05	4.23e-04
Google Inc	7.51e+05	2.18e-04
Sony Corp	2.73e+05	2.17e-04
United Technologies Corp	7.68e+05	1.92e-04
Verizon Communications Inc	7.73e+05	1.67e-04
Siemens AG	6.43e+05	1.49e-04
Honeywell International Inc	3.58e+05	1.11e-04
Safran SA	1.02e+05	1.07e-04
Taiwan Semiconductor Manufacturing Co Ltd	2.25e+05	1.04e-04
ARM Holdings PLC	7.63e+03	5.83e-05
LG Electronics Inc	3.45e+05	5.80e-05

Table 5.1: Firms with the 20 most important contributions to the eigenvector $\left|r^{\lambda_{\min}}\right\rangle$, their respective contributions to the herfindahl, and their reported sales for the year 2015 in thousands of 2015 USD.

Company Name	**Sales (kUSD)**	**Herfindahl Contribution** $\|\|l_i^{\lambda_{\min}}\|\|^4$
Avnet Inc	2.52e+05	8.94e-04
Apple Inc	2.20e+06	6.61e-04
Esprinet SpA	1.59e+04	2.72e-04
Arrow Electronics Inc	1.59e+05	2.39e-04
3M Co	2.50e+05	1.11e-04
Computacenter PLC	2.74e+04	8.06e-05
Data#3 Ltd	1.25e+03	7.86e-05
CDW Corp/DE	1.11e+05	6.33e-05
Emulex Corp	1.26e+03	6.31e-05
Stratasys Ltd	?	6.30e-05
Xilinx Inc	3.04e+04	5.45e-05
JB Hi-Fi Ltd	?	5.19e-05
Entropic Communications Inc	5.68e+02	3.64e-05
Proximus SADP	6.49e+04	3.46e-05
Nuance Communications Inc	2.89e+04	2.97e-05
Premier Farnell Ltd	5.70e+03	2.92e-05
Ingram Micro Inc	3.24e+05	2.86e-05
Systex Corp	6.62e+02	2.83e-05
Northrop Grumman Corp	2.16e+05	2.73e-05
CANCOM SE	8.35e+03	2.67e-05

Table 5.2: Firms with the 20 most important contributions to the eigenvector $\left|l^{\lambda_{\min}}\right\rangle$, their respective contributions to the Herfindahl, and their reported sales for the year 2015 in thousands of 2015 USD.

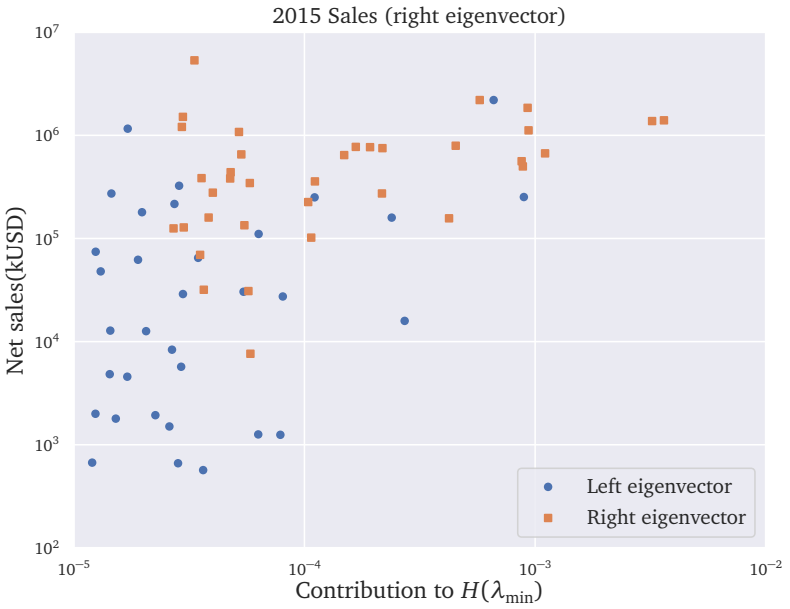


Figure 5.9: Log-log plot of the sales of the 38 main firms contributing to $\left|v^{\lambda_{\min}}\right\rangle$ vs. their contribution to $H(\lambda_{\min})$, suggesting that firms with a higher overlap with $\left|v^{\lambda_{\min}}\right\rangle$ have higher sales.

Plugging this result in the new equilibrium conditions for the N original firms yields:

$$(z_i - \frac{J_{i\star}J_{\star i}}{z_\star})p_i - \sum_{j=1}^{N}\left(J_{ij} + \frac{J_{i\star}J_{\star j}}{z_\star}\right)p_j = V_i + \frac{J_{i\star}V_\star}{z_\star} \tag{5.25}$$

which means that the addition of a firm amounts in effect to *decreasing* all original productivities: $z_i \to z_i - \frac{J_{i\star}J_{\star i}}{z_\star}$ and *increasing* all stoichiometric coefficients: $J_{ij} \to J_{ij} + J_{i\star}J_{\star j}/z_\star$. As clear from Eq. (5.17), this can only decrease the smallest eigenvalue of the matrix $\mathbf{M}_N^\star$ that describes the pre-existing firms with the new firm added. One concludes that a growing economy can only become more unstable with time. This argument is actually closely related to May's original argument about the stability of large ecologies sketched in the introduction of the present Chapter.

In fact, one can show that as the number of links to the most connected node of the network increases, the smallest eigenvalue of $\mathbf{M}$ decreases [80], until the instability threshold is reached. In this case, the fragility of the network comes from the most central hubs, a scenario akin to, but different from, the one of Acemoğlu, Carvalho et al. [4]. This effect might be amplified if firms systematically favour links toward hubs (as suggested in [113]), leading to a "scale-free" input-output network [24]. Interestingly, a stability-constrained growth mechanism for networks, whereby a node is freely added to the network if it does not destabilize the system but induces rewirings in the network until stability is found again if it does, has been found to generate such scale-free networks [211].

The second evolutionary effect is, even for a fixed size N, the complexification of the production process, i. e. , technology progress means that a wider array of products are needed as inputs. If the average productivity z remains the same while the average connectivity c increases, the system eventually reaches the instability point (which in the simplest case reads $z = Jc$). Hence productivity must increase at some minimum rate for the economy to remain stable. But since increasing productivity is costly, one can postulate that the average productivity z will tend to hover around the minimal viable threshold, and sometimes lagging behind, leading to occasional endogenous crises. Similarly, as mentioned after Eq. (5.5), firms tend to optimize their portfolios of suppliers, thereby reducing their redundancy but, by the same token, reducing the effective substitution effects captured by the CES parameter q. As $q \to q_c^+$, the economy will again become unstable.

Finally, the model assumes that firms are perfectly competitive and that equilibrium corresponds to zero profit. Now, in more realistic situations, firms attempt to realize positive profits and distribute dividends. As already noted, if the profit target of firm i is a certain fraction φ_i of its total sales $S_i = z_i \gamma_i p_i$, Eq. (5.11) remains identical but with a decreased effective productivity $z_i \to z_i(1-\varphi_i)$. As firms attempt to maximize their profits, the average effective productivity goes down, until the marginal stability point is reached and a crisis ensues. After the crisis, economic actors revert to more reasonable levels of markups (i. e. reduce φ_i), which makes the economy viable again — until the next crisis.

One could probably come up with other mechanisms that push the economy towards instability, see for example [96, 36, 202]. The conjecture from the present work is that evolutionary and behavioural forces repeatedly drive the economy close to marginal stability. As anticipated by Bak, Chen, Scheinkman, and Woodford [32, 34, 231] and confirmed in this paper, this scenario would be a natural explanation of the broad (Zipf-like) distribution of firm sizes, and of the "small shocks large business cycle" puzzle, that both suggest some kind of criticality. Crises should then be understood as intrinsically non-linear events, where feedback loops of arbitrary size contribute to propagating and amplifying idiosyncratic shocks.

5.5.1 *Introducing dynamics*

There are many directions to explore further. A promising clue is to endow the model with some realistic dynamics, partly along the lines of [52], that would include frictions, myopia, imperfect market clearing, rewiring, etc. This would make the model more realistic, and is a prerequisite to calibration it on empirical data, since within the present static setting crises are signalled by the appearance of negative prices, beyond which the model ceases to make sense. These extensions will now be quickly discussed.

As stressed, the model above only looks at *statics*, while the very nature of fluctuations requires having a full idea of the dynamics that govern

the production of firms. This was the idea behind the model by Long & Plosser, and was picked up again by Bonart and collaborators in [52], and can be summarized by saying that the (linearised) dynamics of a certain quantity x_i (say, the price or the output of a good i) can be represented by an equation such as Eq. (5.1), which we recall can be linearised as

$$\frac{\mathrm{d}\vec{y}}{\mathrm{d}t} \approx \mathbf{A}\vec{y}, \quad \vec{y} = \vec{x} - \vec{x}^{\star}, \tag{5.26}$$

where x_i is a certain quantity, such as the price or the output level of a certain good i, and $x_i^{\star}$ its equilibrium value. η_j is a "random shock", and A_{ij} are the terms of a matrix describing how x_i interacts with other quantities. These interactions can be *within the same firm*, and represent for example the interaction between the output level of a good with its price, or between different firms, such as the interaction between the level of output of a good i and the production and price of one of its required inputs j.

This shall form the basis of a forthcoming paper [101], done in conjunction with Théo Dessertaine, Michael Benzaquen and Jean-Philippe Bouchaud. The main idea behind it is that firms produce each a good with a certain technology set by a production function, that we take to be from the CES family. The inputs that intervene in production, however, were bought at some time in the past, and equivalently at time-point t each firm must forecast how much it will have to produce at $t+1$ to meet its demand, and therefore how many inputs it should buy at that point.

The firms then follow a process of Walras-like *tâtonnement*: firms that are making profit will greedily attempt to increase their output, while those that don't manage to sell all of their production will attempt to reduce it and/or to lower prices to encourage their clients to purchase. These heuristics were strongly inspired by Ref. [52]. The fixed point of the dynamics corresponds therefore to the competitive equilibrium defined in Eq. (5.14), and the remaining question is how the system actually gets there.

When linearising the dynamics, we obtain equations with a similar form to Eq. (5.26), where the variable x_i under consideration can correspond again either to the output or the price of a given good. The terms A_{ij} can also be written in terms of the matrix $\mathbf{M}$ defined above, and, crucially, it is possible to show that the stability condition of that evolution is *also equivalent* to having the matrix $\mathbf{M}$ respect the Hawkins-Simon condition and therefore be an M-matrix.

In this case, the positivity constraint on the variable x_i corresponds also to dynamic stability, and as the system gets closer to violating the Hawkins-Simon condition it also gets closer and closer to dynamic instability, and the time needed for a shock to be dampened-out diverges. Indeed, the solution to Eq. (5.26) can at long times be written as

$$y_i(t) \propto e^{-\alpha_{\min} t}, \tag{5.27}$$

with $-\alpha_{\min}$ the smallest eigenvalue of the stability matrix $\mathbf{A}$. The system is stable if $\alpha_{\min} > 0$, and one can show that $\alpha_{\min}$ is linked to $\lambda_{\min}$, the smallest

eigenvalue of the matrix $\mathbf{M}$ of Sec.5.1, and that $\alpha_{\min} \to 0$ as $\lambda_{\min} \to 0$. This scenario therefore corresponds to a situation where a shock to the system is amplified through the supply-chain very strongly, and therefore lingers on a very long time before being absorbed.

5.5.2 *Perspectives*

The models introduced above provide theoretical clues on possible channels of propagation and amplification of fluctuations in large economies. The introduction of granularity showed that the heterogeneity of firm sizes is responsible for a relevant portion of observed macroeconomic fluctuations, and the following discussion on network models proved the necessity to take into account the different degrees at which different firms can be interconnected.

They constitute, in short, models that try to go beyond the idea of independent shocks affecting firms. In a way, it is a lot like the advances made after Louis Bachelier's initial model for stock prices fluctuations as geometric brownian motion [28], where prices p_t are modelled to evolve as

$$\ln(p_{t+1}) - \ln(p_t) = m + \sigma\eta_t := r_t, \tag{5.28}$$

where r_t is called the *return* and encodes the random fluctuations of the price, it has an average of m, and a square-root volatility σ. The last term η_t is a random variable that is taken as a gaussian in Bachelier's model.

Despite pervasive usage in financial mathematics and in certain classrooms, this model is empirically found to be too simple, and is unable to explain a series of empirical facts. A good review can be found on Ref. [93], but some of the interesting features found in return time series is their *fat-tailed* nature (meaning that the variable η shouldn't be taken as gaussian) and their intermittency (meaning that, from time to time, there are bursts of high volatility in return time series). I've illustrated this using recent financial data from the S&P500 stock index on Figure 5.10.

Promising work has been done in studying and explaining these statistical facts, and how they can emerge from the microscopic actions of agents interacting in financial markets [66, 67]. An interesting avenue has been proposed recently in [152, 15], where the authors propose to model the prices of financial assets as evolving via coupled dynamics, similar to the form argued in Eq. (5.1) with the addition of stochastic noise. Their method allows for analytical and numerical results, showing therefore that the addition of coupling between different assets with non-linear feedback loops allows to recover features like the bursts of volatility observed in Figure 5.10, and to reproduce some of the statistics of the returns r_t of Eq. (5.28). Similar results have been obtained in the study of business cycles in Ref. [207].

Finding similar mechanisms in the *fundamental* variables describing a firm, i. e. those linked to its production, sales or other non market-related variables, would be ideal to gain a better understanding of large macroeconomic fluctuations. In finance, however, many such results rest upon a well documented base of statistical stylized facts that act as constraints that any realistic model should reproduce. The same has been attempted with the statistics of the *growth-rates* that describe the dynamics of firm growth (see Refs. [251, 164, 58] and others in Chapter 6). The goal of the following section is to explore these models and to understand the nature of the statistical stylized facts that describe company growth.

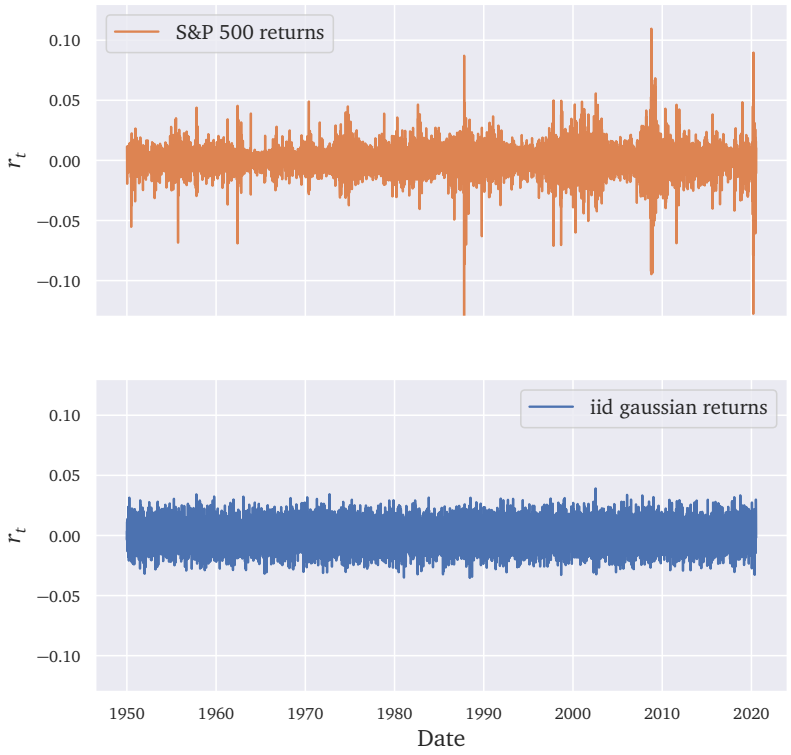


Figure 5.10: Top: daily returns r_t associated to the S&P500 stock index from 1950 until the 11$^{\text{th}}$ of July 2020. Bottom: synthetic gaussian iid returns calibrated to have the same mean and variance as the S&P500 returns. The real returns time series exhibits a series of statistical properties: in addition to large fluctuations that one can specifically pin to certain events (1987's Black Monday, 2020's Covid-19 crisis) it is clear that there are periods of turbulence, when the market is very volatile. This last point can be seen by comparing the fluctuating envelope of the top curve, compared to the relatively flat envelope on the bottom. The data for this curve was downloaded on 11/07/2020 from Yahoo! Finance.

Part III

EMPIRICAL PROPERTIES OF FIRM GROWTH

The goal of the following Chapter is to show the main statistical facts governing firm growth.

Now that the reader is familiar with power-law distributions, their presence in the description of firm sizes and the impact they may have on the *small-shocks, large business cycle* puzzle, I will give an empirical framework to the description of firms' *growth-rates*.

6

EMPIRICAL ANALYSIS OF FIRM GROWTH

6.1 INTRODUCTION: GIBRAT'S LAW

An established empirical fact in the description of firms, as shown in Section 4.1 and in particular in Figures 4.2 and 4.3, is that their "size", be it their sales or their market capitalization for example, is well described by a Zipf law, viz. a power-law with exponent $\mu \approx 1$. This was noticed already in 1931 by Robert Gibrat [128], with Simon [244] noting that this is not what one could expect from, for example, a notion of an "optimal firm size" where firms try to maximize profits and minimize operating costs.

Gibrat proposed instead a *stochastic growth model*, based on multiplicative growth which, as I showed in Sec. 3.3, can lead to power-law distributions after adding in very mild assumptions. This was known as Gibrat's *law of proportionate effect*, and its main claim is as follows.

For a certain quantity x that can be used to describe a firm, such as its market capitalisation, net sales or total assets,[1] call $x_i(t)$ its value for firm i at time t. Then Gibrat's multiplicative growth model states that

$$x_i(t+\Delta t) = \left(1 + g^i_{x,\Delta t}(t)\right) x_i(t) \tag{6.1}$$

where $g^i_{x,\Delta t}$ is defined as the (percentage) growth rate of the quantity x_i between t and $t+\Delta t$. When this rate g is small compared to 1, as is usually the case, we can identify it to the *log-growth rate*,

$$g^i_{x,\Delta t}(t) = \log\left(\frac{x_i(t+\Delta t)}{x_i(t)}\right), \tag{6.2}$$

which is the usual way the growth rate is defined, and the definition I use henceforth.

The strong version of Gibrat's law then claims that the statistics of g are independent of x, or in other words that

$$p(g|x) = p(g), \tag{6.3}$$

i. e. that the distribution of g conditional on x is just the distribution of x. A considerable body of evidence, however, contradicts this hypothesis (see [12] and references therein). It is found, for example, that the standard deviation of g depends on x, as

$$\text{std.}(g|x) := \sigma(x) \propto x^{-\beta}. \tag{6.4}$$

1 The precise quantity one uses, and where it is a "flow" variable or a "stock" variable, does not change the analysis, see Ref. [20].

Much has been also said of their distribution $p(g)$ [251, 12], with a large body of research advocating for a Laplace, Subbotin or similar "tent-shaped" distributions for its description [57, 58, 56, 55]. Regarding the tails of the distribution, it has also been proposed that they are well fit by power-laws – even Cauchy-like [274] – or stretched exponentials.

It has also been suggested that these properties hold when one divides firms in groups of different "sizes" (meaning for similar values of the underlying variable x) and then proceeds to centring and rescaling (subtracting the mean and dividing by the standard deviation) their growth rates [164]. These observations have led to departures from the original Gibrat's model from Eq. (6.1), where the g variables were modelled as random numbers with a size-independent distribution.

In this Chapter, I will attempt to give a description of these growth rates as a way of understanding the dynamics behind the rise and fall of firms. The inspiration for the present work is the empirical work that has been done in the past decades in finance (see e. g. [66, 67]), where modelling has been successful in explaining many statistical phenomena.

The main difference, as the reader shall see, lies in the crucial distinction that financial data is plentiful, and one usually has very long time series for each financial asset. In contrast, firm-level data consists often of relatively short time-series (a few tens of data points at the most) that one has to manipulate carefully. How to handle this properly is at the heart of this Chapter, and taking into account the heterogeneous nature of volatility is, I think, my main contribution to this area. The following work has not yet been published, and was done in conjunction with Angelo Secchi and Jean-Philippe Bouchaud.

6.1.1 *Description of the dataset*

In the following I will focus on the dynamics of fundamental and financial variables describing firms. To this end, I use four different data sources. The first dataset consists of the CompuStat/CRSP database, where I've kept only firms whose stocks are traded in the NASDAQ and NYSE exchanges [2], with a timespan ranging from 1962 to 2013 and corresponding to 20,140 firms sampled with a quarterly frequency. This quarterly sampling led me to use this as the main dataset for this study.

The second dataset is the French National Institute of Statistics and Economic Studies' (INSEE) FICUS dataset, ranging from 1994 to 2007 and containing all legally registered French "firms" in that time period, barring those subject to confidentiality constraints because of, e.g., activities within the military sector. The definition of a firm in this dataset is very broad and corresponds to the French "SIREN" system. It is an unique 9 digit code used to identify French businesses that does not change through their history.

2 All quantities have been converted to their USD value using FOREX data.

The main characteristic of this is that different units with distinct SIREN numbers may actually be part of a single firm [3].

Starting from an initial pool of around 6 million firms, I have selected only those that were present every year from 1994 to 2007 and that declared having at least one employee during that period. In all, this subset amounts to 374,714 firms sampled with a yearly frequency from 1994 to 2007.

I should also note that the INSEE dataset consists of entries that are filled by the firms themselves. They are therefore prone to containing mistakes, and this supposed a preliminary work of data cleaning that lasted a few months to ensure that no aberrant entries were present in my working dataset. These consisted most of the time of misplaced digits, in which case the entry was discarded after checking that the value entered was many orders of magnitude away from what would be expected, or of dates that were entered in the wrong format, in which case the entry was systematically fixed. This was done using the *Centre d'accès sécurisé aux données* (CASD) system to grant access to confidential data produced by French public entities, under project reference REGSTAT.

The third and fourth datasets consist of the FactSet Supply Chain Relationships database and the Reuters Fundamentals dataset. The former is a list of relational data between firms, stating if firms A and B have a client/supplier relation. It is built by collecting information from primary public sources such as SEC 10-K annual filings, investor presentations and press releases, and covers about 23,000 publicly traded companies with over 325,000 relationships. The Reuters Fundamentals dataset allows us to match the supply chain data with the fundamental and financial variables of each firm.

Using the FactSet database, I've built a supply chain graph as in Sec. 5.4 using the relational data spanning from 2010 to 2019 and selected the largest weakly connected component[4] within that network, consisting of 12,257 firms which I've then matched to the Reuters Fundamentals database. This last database has a sampling rate of one data point per year for each firm.

6.2 STUDY OF THE GROWTH RATES

In the following, I shall study the growth rate of a quantity x as defined by Eq. (6.2). The smallest sampling frequency corresponds to that of the CRSP/Compustat dataset, and corresponds to $\Delta t = 1$ quarter. For simplicity, I will therefore set the base unit of time as $\Delta t = 1$ q., with e. g. $g^i_{x,4}(t)$ the growth rate corresponding to $x_i(t)$ over one year.

As a preliminary, I have plotted the yearly growth rate distribution for the net sales of firms, $g_{x,4}$ for with x the net sales, for both the CRSP/COMPU-

3 For example, Total SA, Total Holdings SAS SNEA and Total Marketing correspond all to distinct SIREN identifiers and are therefore considered as distinct units in this dataset, despite all being part of the same company.

4 A weakly linked sub-graph is such that any two nodes in it can be linked by a directed path.

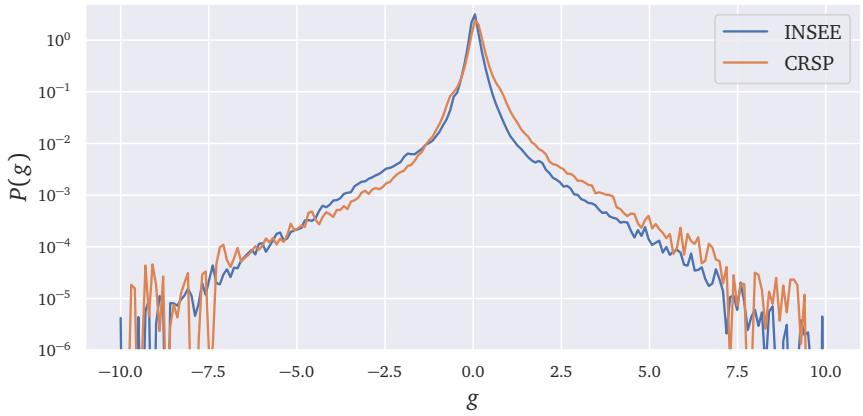


Figure 6.1: Yearly growth rate distribution for firms in the INSEE and CRSP/COMPUSTAT databases, plotted using a gaussian kernel estimation of their density. The growth rates have been all pooled together and then centred and rescaled using the global mean and standard deviation.

STAT and INSEE datasets on Figure 6.1. It is apparent that the distributions show a Laplace-like "cusp" in the central region, and fat-tails that resemble those of e.g. the Subbotin distribution of Ref. [55]. A striking feature however is that this distribution also appears when one considers growth rates with $\Delta t = 1$ q. or $\Delta t = 8$ q., something that doesn't conform to the intuition one may get from the Laplace distribution, as I sketch below.

6.2.1 *A critical discussion*

From the definition given in Eq. (6.2), it is clear that yearly growth rates should be statistically equivalent to the sum of four quarterly growth rates. This is a direct consequence of the relation $g^i_{x,4}(t) = \sum_{k=0}^{3} g^i_{x,1}(t+k)$.

My argument is that, given this fact, the Laplace-like "cusp" observed in previous references is an artefact and gives a wrong picture of firm dynamics. As shown in Appendix A.2, it is easy to show for example that if g is the sum of two Laplace random variables, rescaled to have a variance of one, then its density reads

$$p_{L,2}(g) = \frac{1}{2}(1+2|g|)e^{-2|g|} \underset{g \ll 1}{\approx} \frac{e^{-2g^2}}{2}, \tag{6.5}$$

showing that the cusp is not stable by aggregation. It is also possible to show that this also holds when the two Laplace increments are correlated positively or negatively, see Appendix A.2.2. This, to me, is a hint that previous studies may have been using the wrong definition for g.

A way of understanding why this may have been the case becomes clear with the following thought experiment. When one has in mind Eq. (6.2) and the modelling of $g^i_{x,\Delta t}$ as a random variable, one is actually thinking of a purely hypothetical experiment where one would be able to observe the

evolution of firm i between t and Δt and re-run this a very large number of times, for the *same firm*.

One would then be able to compute explicitly the distribution of $g^i_{x,\Delta t}$, i. e. the growth rate distribution *for firm i only*, and this distribution would describe the idiosyncratic dynamics of the variable x_i between t and Δt. Were one to look then into larger values of Δt one should then see a convergence to a gaussian distribution in the central region, even if it is slow, as one observes for e. g. prices of stocks of these companies [218].

The question that arises next is what could explain the prevalence of these *tent-shaped* distributions, and where the previous analyses done may have gone wrong.

6.2.2 *Explanations of past results*

How to explain the persistence of the tent-shaped distribution in the literature? If indeed Eq. (6.2) is taken as the base model, then as stressed before, larger values of Δt should result in a distribution with a gaussian central region. Since this is not the case, one of the remaining possibilities is that we are, so to say, mixing apples and oranges when plotting the distribution of the growth rates of *all* firms together.

Past work has shown for example that the volatility of growth rates, measured by their standard deviation or any other proxy denoted σ, should depend on the "size" of the firm, as stressed in Eq. (6.4). A very naive question that arises once one is aware of this is if the distributions observed in past literature, and possibly some of the fat tails claimed in the growth rate distribution, aren't just a consequence of mixing distributions.

The first approach to this is looking at the different values of σ in the data. As I am dealing with very short time series (each firm in the INSEE data only gives 12 data points, for example), I define it here to be given by

$$\sigma^i_{x,\Delta t} = \sqrt{\frac{\pi}{2}}\mathbb{E}_t\left\{\left|g^i_{x,\Delta t}(t) - \mathbb{E}_{t'}\left[\left|g^i_{x,\Delta t}(t')\right|\right]\right|\right\}, \tag{6.6}$$

a less noisy estimator of the standard deviation than the square-root of the empirical variance.[5]

I computed the volatility of each firm in the INSEE and CRSP/Compustat dataset using the above Eq. (6.6) and plotted the results on Figure 6.2. Surprisingly, the distributions are well fit by inverse gamma distributions of density

$$p_\sigma(\sigma) \propto \exp\left(-\frac{b}{(\sigma - a)}\right)\sigma^{-1-\mu}\mathbf{1}_{\sigma>0}, \tag{6.7}$$

where a and b are parameters that control the shape of the distribution for $\sigma \ll 1$ and μ is the tail parameter of the distribution, $p_\sigma(\sigma) \underset{\sigma\gg 1}{\propto} \sigma^{-1-\mu}$.

5 Indeed, this formula converges to the standard deviation when g is a gaussian random variable.

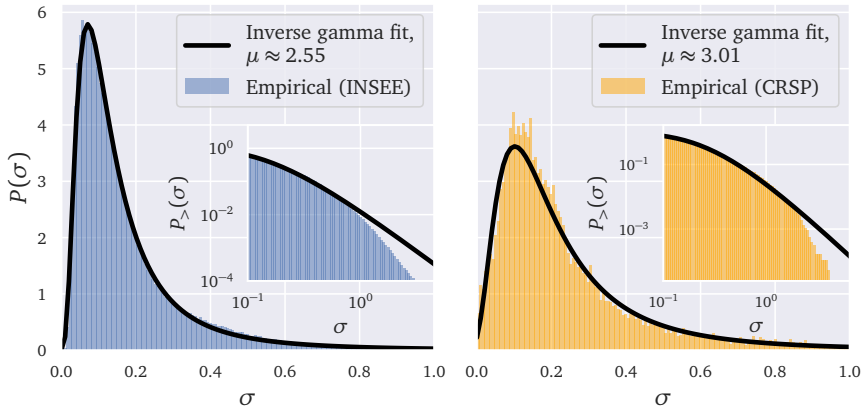


Figure 6.2: Empirical histograms of the sales volatilities for the INSEE and CRSP/-CompuStat datasets. Inset: survival function $P_>(\sigma) = \int_\sigma^\infty \mathrm{d}\sigma'\, p(\sigma)$.

This has strong implications. For example, a number of researchers have argued in favour of growth-rate distributions with tail exponents $\mu \approx 3$, but one could argue instead that this is the result of mixing growth rates with heterogeneous volatilities, as is the appearance of the tent-shaped cusp in the central region.

Indeed, if one considers a random variable $g = \sigma\tilde{g}$ that is the product of a gaussian random variable $\tilde{g}$ with a random variable σ with a density that behaves asymptotically as $p_\sigma(\sigma) \underset{\sigma\gg 1}{\approx} \sigma^{-1-\mu}$, then an elementary computation for the probability density of g (for large g) yields:

$$p_g(g = \sigma\tilde{g}) = \int \frac{\mathrm{d}\tilde{g}}{\tilde{g}}\, p_{\tilde{g}}(\tilde{g}) p_\sigma\left(\frac{g}{\tilde{g}}\right) \underset{g\gg 1}{\propto} g^{-1-\mu} \int_{-\infty}^{\frac{g}{\sigma_m}} \mathrm{d}\tilde{g}\, p_{\tilde{g}}(\tilde{g})\tilde{g}^\mu, \quad (6.8)$$

where $p_{\tilde{g}}$ is the distribution of a gaussian random variable. This means that mixing gaussian distributions with heterogeneous volatilities gives a random variable that inherits the tails of the volatility distribution.

As a stronger way to push this argument I have also run the following experiment. Take the set of the σ_x^i volatilities for the sales on the CRSP/-Compustat dataset, and construct a synthetic random variable by drawing a number from a normal distribution and multiplying it by a randomly chosen value of σ_x^i. The result of this numerical experiment is visible in Figure 6.3, showing that they are virtually indistinguishable from the observed unscaled/mixed growth rates from the CRSP/Compustat dataset. Besides having roughly the same tail, the distribution also features a cusp in the central region, something which can now be interpreted as coming from firms with very low growth rate volatilities. This simple explanation should be contrasted with the complex mechanisms used thus far to explain the Laplace central region with heavy tails in the existing literature [57, 58, 72, 115], that rely mostly on self-reinforcing mechanisms.

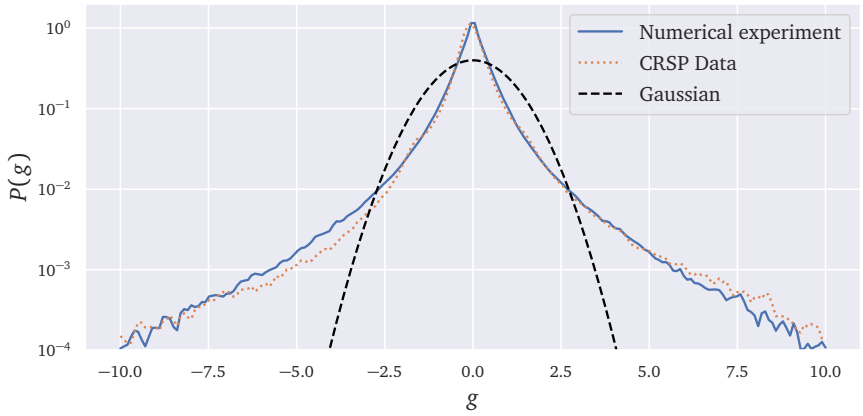


Figure 6.3: Empirical density plot of the random variable defined as the multiplication of a random gaussian number with a volatility randomly drawn from the set of CRSP/Compustat net sales yearly growth volatilities, along with the empirical density of the unscaled CRSP/Compustat growth rates. Both densities were estimated using a gaussian kernel. The dashed line represents a gaussian distribution for comparison.

Having now shown that it is necessary to take heterogeneity, or more precisely heteroskedasticity, into account, I can now turn to a more in-depth study of the statistical facts behind growth rates.

6.3 STATISTICAL FACTS ON GROWTH RATES

Given the heterogeneous nature of volatility that we have shown, there is no *à priori* reason to think that the distribution of growth rates should be the same across firms, nor indeed across time periods. One can nevertheless make the sensible hypothesis that this distribution may only depend on intrinsic characteristics of the firm, such as its available technology, sector of activity, age or any other relevant quantity that describes it. As these are all things that one expects not to vary significantly during the history of a given firm, I make the hypothesis that the growth rate distribution *for a given firm* is time independent, but retain the possibility that two different firms have different corresponding distributions.

Hence, I have started by reproducing known results concerning the statistics of growth rates. The simplest among these consist of looking at averages of different quantities within "bins" of firm age, size or firm sectors. For a certain category $\mathscr{C}$ of firms, I define the average growth rate and volatility of quantity x with that category as

$$\mathbb{E}_{\mathscr{C}}(g_{x,\Delta t}) = \mathbb{E}_{t,i\in\mathscr{C}}\left[g^i_{x,\Delta t}(t)\right], \quad \sigma_{\mathscr{C},\Delta t} = \mathrm{std}_{t,i\in\mathscr{C}}\left(g^i_{x,\Delta t}\right). \quad (6.9)$$

Unless specified otherwise, in what follows I take $\Delta t = 4$ q. $= 1$ year.

As a first example, I have binned firms in the INSEE dataset by their age in years, so that categories $\mathscr{C}$ correspond to sets of firms with the same age

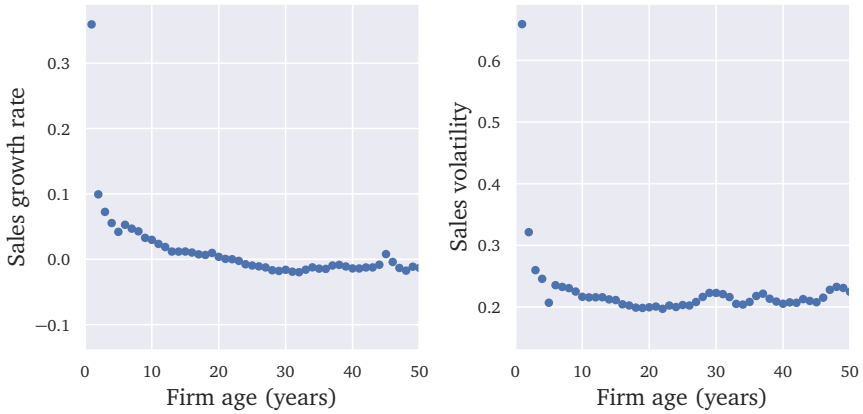


Figure 6.4: Net sales average growth and volatility of firms in the INSEE dataset.

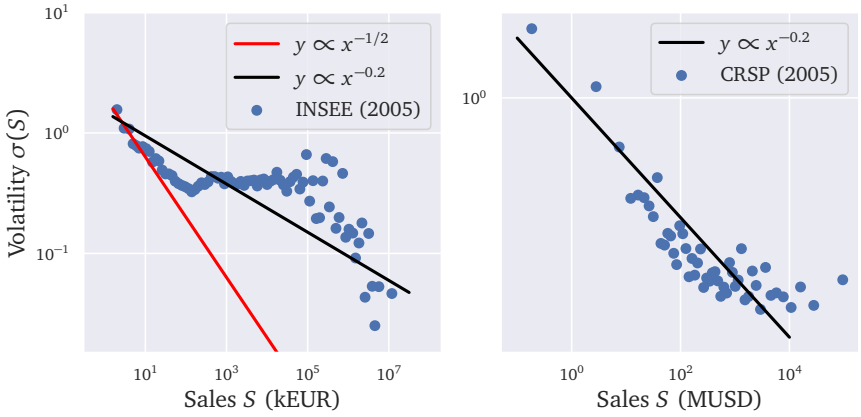


Figure 6.5: Scaling of sales growth volatility $\sigma(S)$ vs. sales S for the year 2005, showing a clear scaling $\sigma(S) \propto S^{-\beta}$ with β close to 0.2 for the CRSP dataset. The INSEE dataset does not show such a clear scaling relation

after their creation. I have plotted the average growth rate and volatilities of the net sales vs. firm age in Figure 6.4. My findings roughly confirm known results [77, 170], namely that "young firms grow fast". They also tend to be more volatile, but this only concerns *very* young firms. I must also stress that this is an expected result: the very fact that I can compute growth rates as defined in Eq. (6.2) implies that these firms have survived, and so that there is a survivor bias for small, young firms in that the average is necessarily taken over survivors only.

I've also binned firms by their "size", defined as their net sales on Figure 6.5 (for the INSEE and CRSP/Compustat datasets) or their market capitalisation on Figure 6.6 (for CRSP/Compustat only), and computed the volatility of the growth of x, where $x = S$ for the net sales or $x = M$ for the market cap., and found that a scaling $\sigma(x) \propto x^{-\beta}$ with $\beta \approx 0.2$ holds rather well for the CRSP data.

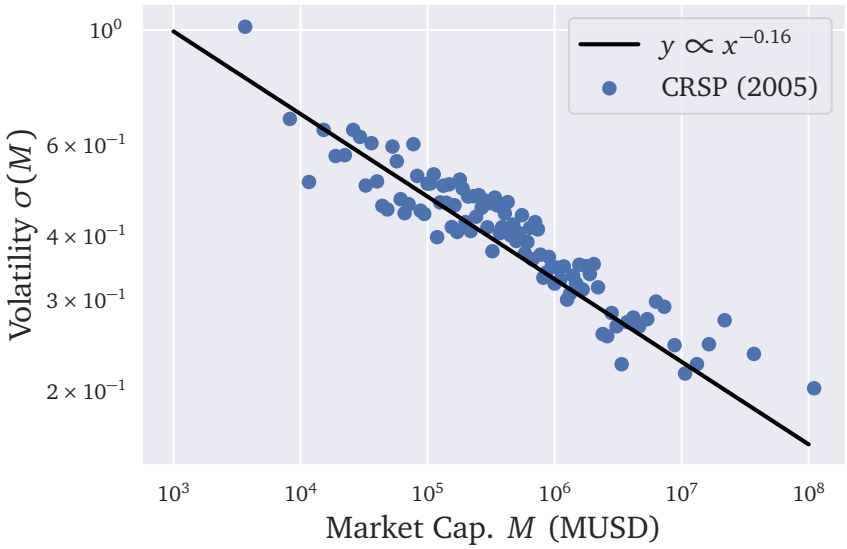


Figure 6.6: Scaling of market cap. growth volatility vs. market cap. in the CRSP dataset in the year 2005. Note that the scaling is similar as that in Figure 6.5.

In contrast, in the INSEE data the scaling doesn't hold as well, and one sees that a scaling with β closer to $1/2$ seems to hold in the beginning but breaks down afterwards.

An explanation to this may be attempted by saying that the base units for the INSEE database are not *real firms*, in the sense that an entity identified by a SIREN is not always an independent firm, but may be part of a larger firm made up of many SIREN units. In contrast, the scaling of β close to $1/2$ for firms with small sales seems to be in line with considering that units are mostly very small independent firms, and are therefore subject to standard CLT-style fluctuations with an exponent that corresponds to the square root of the size, an argument close to the one sketched in Chapter 4. See also below for a more thorough discussion on this and an explanation for $\beta \approx 0.2$ for CRSP.

In the same vein, I now take the firms' Bloomberg sectors for categorizing them. Figure 6.7 shows the average volatility of the yearly sales and market cap. growth rates for different sectors. It's easy to see that some sectors are much more volatile than others, and that volatility in the sales seems to correlate with volatility in the stock price.

To continue this study of the relation between sales or market cap. volatilities and other factors describing a firm I have also looked into the FactSet/Reuters dataset. I first check for the intuition developed previously, namely that volatility in sales and market cap. growth rates seem correlated, by plotting a two-dimensional hex-bin plot of these two quantities. Clearly, firms that have very volatile sales also tend to be rather turbulent in the markets, as one would expect intuitively.

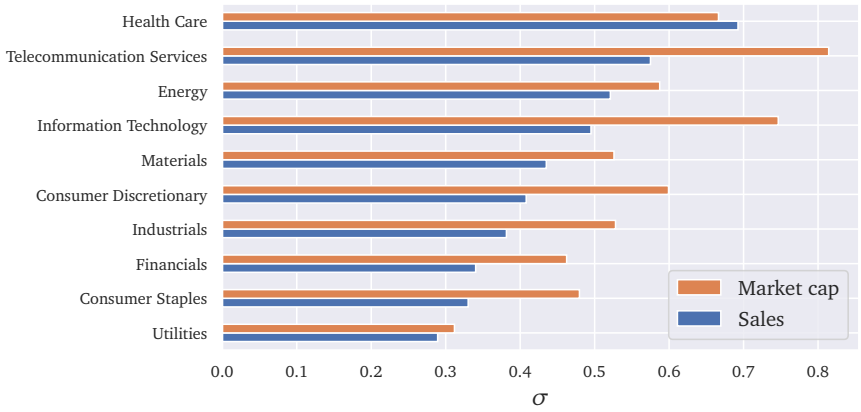


Figure 6.7: Sales and market capitalisation volatility by sector for firms in the CRSP/Compustat dataset.

In addition to this, I have attempted to see how volatility depends on some network-related quantities to the firms. The rationale behind this is close to the arguments given in Sec. 5.5, and relies on the notion that firms need to buy from one another for their production.

A very stylized model for this, corresponding roughly to the description in Sec. 5.5.1, can be described as follows. Call $\pi_i(t)$ the production output of firm i at time t. Naively, one can say that

$$\pi_i(t) = f_i\Big(\{Q_{ij}(t-\Delta t)\}_{j\in\mathcal{N}(i)}\Big), \tag{6.10}$$

where $Q_{ij}(t-\Delta t)$ is the input that firm i bought from a supplier j at some point in the past $t-\Delta t$ and $\mathcal{N}(i)$ is the set of suppliers of i (the firms from which it gets its inputs). As before, the production function f_i describes firm i by taking into account its technology and the degree at which various inputs intervene in its production.

One may then roughly expect that firms that lie in central positions in the network, that is those with a large number of clients and suppliers, should be less volatile, as the different fluctuations in the Q_{ij} quantities (the "supply shocks") and in the demand for its production may average out.

To test this, I've taken computed the Eigenvalue Centrality (also known as PageRank) of the firms in the FactSet dataset between 2013 and 2019. As before, I've made two-dimensional hex-bin plots comparing these quantities to the volatility of firms, as defined in Eq. (6.6), and to the average rescaled sales. This last quantity consists of the average of $s^i(t)/\text{Med.}_i(t)$ over t, that is the average of the sales of a firm divided by the median sales of all other firms within the same year.

In line with what I've said in Sec. 5.4 and with Refs. [191, 4], firms tend to be "larger" in terms of sales, and also to be less volatile. It is not clear, however, what the causality relation may be (e. g. if more central firms are larger than therefore less volatile, as made explicit previously, without establishing any clear link between volatility and centrality). I

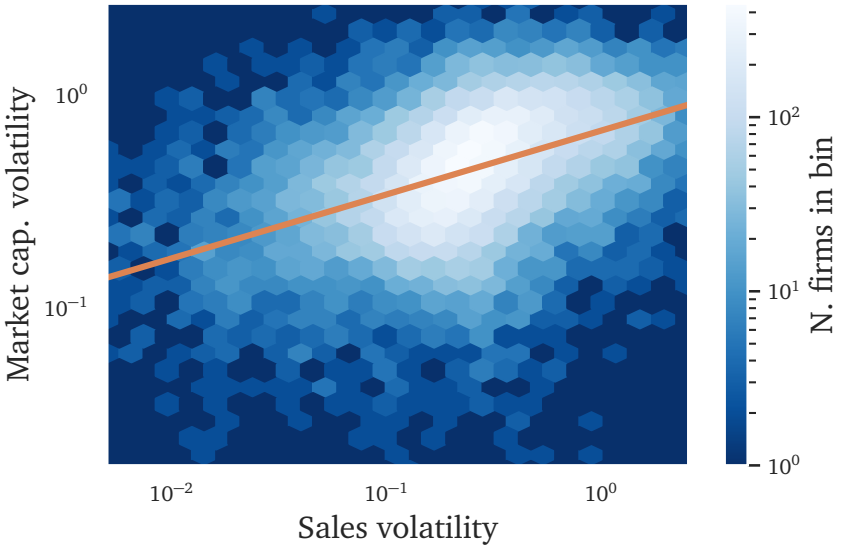


Figure 6.8: Two-dimensional hex-bin plot of the sales growth rate volatility vs. market cap. growth rate volatility for the Reuters/FactSet dataset, along with a simple least squares fit $\log(y) = a\log(x) + b$.

stress however that this result is only shown for illustrative purposes, as I have not yet found that the link between network position and volatility is statistically significant.

6.3.1 *The size-volatility scaling relation*

In this subsection, I provide empirical evidence that the $\sigma(x) \propto x^{-\beta}$ with $\beta \approx 0.2$ scaling relationship stems from the internal structure of firms. The original findings by Stanley et al. [251, 12, 72] considered the "size" of a company defined by its sales, and showed that such a scaling relation extends when one considers countries with x as their GDP.

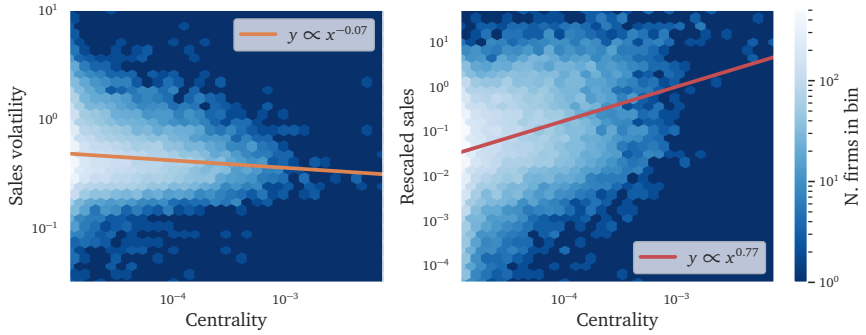


Figure 6.9: Eigenvalue centrality vs. volatility and (rescaled) sales, along with simple least squares fits $\log(y) = a\log(x) + b$. Note that more central firms tend both to be less volatile and to have larger sales.

That larger firms fluctuate less is not contrary to intuition, since large companies are likely to be more diversified. If one expected, however, a firm of size S to me composed on N independent sub-units of size s with $Ns = S$, then quite clearly its fluctuations would scale as $\sigma(S) \propto S^{-1/2}$ because of the central limit theorem, as explained in Chapter 4. An explanation to deviations from this was proposed in Ref. [72], where the authors argue that correlations between the sub-units that make up a firm could be responsible for the slower decrease of volatility as a firm gets larger.

Sutton [254], however, argues that the value $\beta \approx 0.2$ is very stable across time, and that this is at odds with the view that the exponent has an origin in the correlations of company sub-units. Instead, he proposes a mechanism where this behaviour arises from the structure of the firm: he decomposes a firm in N sub-units with sizes determined by a maximum entropy principle[6] and deduces an exponent $\beta = 1/4$.

This "micro-canonical" approach isn't scale invariant, and a similar approach was proposed by Wyart and Bouchaud [279], where they propose that firms may themselves be composed of a number of independent sub-units stemming from a power-law distribution, with sizes that are also power-law distributed. This formulation has the advantage of being "scale-invariant", in the sense that if one takes two firms with such structure and "fuses" them into a bigger firm then the basic structure is preserved. They find analytically that the exponent β comes directly from the exponent of the power-law distribution of sub-unit sizes, showing a similar origin to the anomalous decrease of volatility as Gabaix' granularity [120]. Here, I provide evidence that this mechanism is a plausible explanation for the $\sigma(x) \propto x^{-\beta}$ scaling relationship found in the data, and therefore also an explanation for the fact that the same scaling relation can be extended to explain the scaling relationship between countries' GDP and its growth volatility found by Stanley and collaborators.

6.3.1.1 *Artificial clustering*

In order to test this hypothesis of scale invariance, I consider all 20,140 firms in the CRSP/Compustat dataset and put them into "clusters" or "super firms" containing 12 firms each. Not all firms are present during the whole span of the data, but this in my view mimics the fact that the number of sub-units in each firm may itself vary in time, through normal firm growth processes such as mergers, acquisitions or the founding of new internal structures.

I next define the net sales of each cluster as the sum of the sales of its components for each year. In order to gather more statistics, I "deflate" these sales by dividing them by the cross-sectional median in the year. In other words, for cluster α I define its sales for year t $S_\alpha(t)$ as:

$$S_\alpha(t) = \sum_{i \in \alpha} S_i(t), \tag{6.11}$$

6 In other words, he takes the size of the company S and considers the most probable way to partition it into N sub-units with different sizes

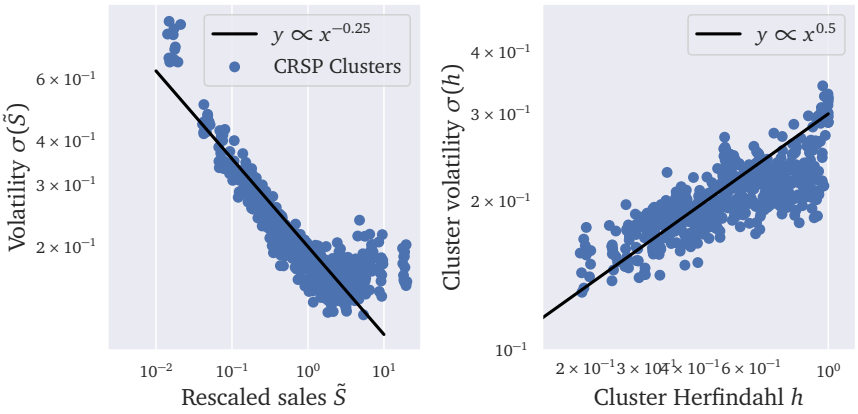


Figure 6.10: Results of the aggregation experiment. Note that the $\sigma(s) \propto s^{-\beta}$ scaling relation is preserved by the aggregation procedure, and that there seems to be a relationship between volatility and the herfindahl index scaling as $\sigma(h) \propto h^{1/2}$ as one would expect from Eq. (4.5) as stated in Sec. 4.1.

where S_i is the net sales of firm i. The rescaled, or deflated, sales now read

$$s_\alpha(t) = S_\alpha(t)/\text{Med.}\left(S_{\alpha'}(t)\right). \tag{6.12}$$

I can then compute the yearly growth rate of the cluster's sales, as

$$g_S^\alpha(t) = \log\left(\frac{S_\alpha(t+1)}{S_\alpha(t)}\right). \tag{6.13}$$

Finally, I bin $g_S^\alpha(t)$ conditional on the value of $s^\alpha(t)$ and compute the standard deviation of the growth rates, so that I am computing the volatility of growth rates conditioned on the clusters' rescaled sales. According to the Wyart-Bouchaud model, I also expect a scaling relation of the form $\sigma(s) \propto s^{-\beta}$.

This aggregation carries the advantage of allowing us to see some information on the structure of the cluster's we've built. It is possible, for example, to compute the herfindahl index of the cluster at time t, defined as

$$h^\alpha(t) = \frac{\sum_{i\in\alpha} S^i(t)^2}{\left(\sum_{i\in\alpha} S^i(t)\right)^2}, \tag{6.14}$$

and equivalent to the definition I used in Sec. 2.3.1.

I've run this synthetic aggregation of firm experiment 10 times and shown all of the results on Figure 6.10. As one sees the aggregation procedure preserves the scaling relationship, with the additional feature that clusters with a higher herfindahl tend to be more volatile.

I now compare this with a numerical experiment on the Bouchaud-Wyart model. The procedure is rather simple, I take $N = 10^7$ units with sizes

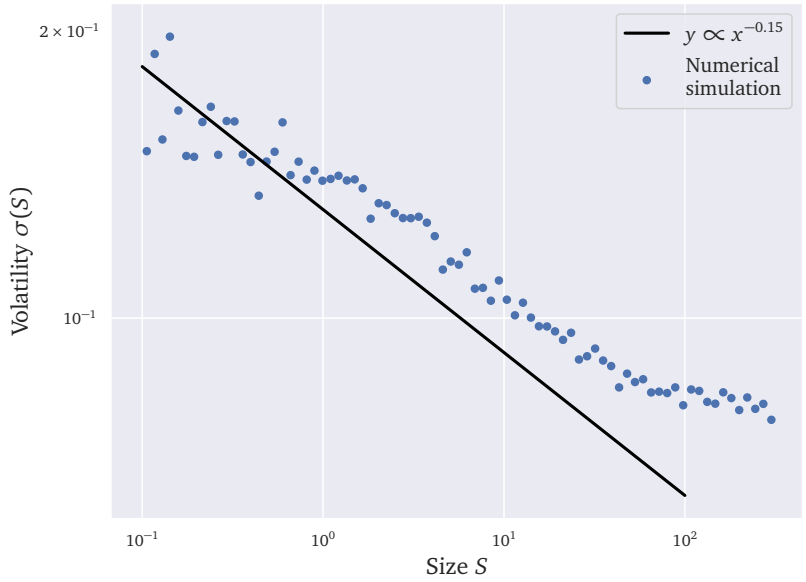


Figure 6.11: Results of the completely synthetic aggregation experiment. Note that these are qualitatively the *same* results as those in Figure 6.10.

s_i distributed as $p_s(S) \propto s^{-1-\mu}$ for large s. I've next grouped them into $N_c = 2 \cdot 10^4$ clusters, with cluster α made of K^α units and where K^α is power-law distributed, with $p_K(K) \propto s^{-1-\nu}$ for large K and with $\nu > \mu$. The initial size of cluster α is then given by

$$S_0^\alpha = \sum_{i \in \alpha} s_0^i. \tag{6.15}$$

I then allow each sub-unit to grow multiplicatively by a random gaussian factor, so that the size of the cluster after one iteration is given by

$$S_1^\alpha = \sum_{i \in \alpha} s_1^i = \sum_{i \in \alpha} s_0^i (1 + \sigma \eta_i), \tag{6.16}$$

where the η_i variables are centred gaussian variables with unit variance.

I can then compute the growth rate volatility conditional on the cluster size and herfindahl, and show the results on Figure 6.11. Note that these are qualitatively the same results as those of Figure 6.10, showing that the scale invariance of the number of firm sub-units and their respective sizes is a plausible explanation for the observed empirical results.

After providing empirical validation for the Bouchaud-Wyart scenario, which also implies a plausible explanation for the scaling of countries' GDP fluctuations, one should remark that a prediction from that model is that the size growth rate conditional on size should become gaussian at least for entities with large sizes, something that until now is at odds with empirical findings. As the reader will later see, firm growth rates do in fact possess a gaussian central region.

6.4 GROWTH RATES ARE APPROXIMATELY GAUSSIAN

After this careful study of volatility, and after having shown that the reason behind the tent-shaped distribution and the fat-tails of firm growth rates is the heterogeneity of their volatilities, I now turn to the distribution of the rescaled growth rates $\tilde{g}$. These are defined using a common technique to rescale time series with heterogeneous volatilities [66], also named "leave one out" rescaling, viz.

$$\tilde{g}^i_{x,\Delta t}(t) = \frac{g^i_{x,\Delta t}(t) - \mathbb{E}_{t'}\left[g^i_{x,\Delta t}(t')\right]}{\sigma_{t'\neq t}\left[g^i_{x,\Delta t}(t')\right]}, \tag{6.17}$$

where $\sigma_{t'\neq t}$ corresponds to Eq. (6.6) but where the observation for $t' = t$ is left out. This is similar to the view adopted to study the statistics of stock price variations in [134], and comes from writing

$$g^i_{x,\Delta t}(t) = \mathbb{E}_{t'}\left[g^i_{x,\Delta t}(t')\right] + \sigma^i_{x,\Delta t}\tilde{g}^i_{x,\Delta t}(t). \tag{6.18}$$

I stress again that the main difference between this approach and that adopted in previous research is that I rescale by each firm's *idiosyncratic* volatility to study the statistics of growth rate, following the procedure used in [134] to study the statistics of stock price variations.

Going back to the thought experiment proposed in Sec. 6.2.1, I believe in fact that if one were to look at the growth-rate distribution of a *single firm* then it is likely that one would see a gaussian distribution (at least for large Δt!) with a width that depends on the complete characteristics of the firm, and not only on its size. Accordingly, I've computed the variable $\tilde{g}$ for the following different cases:

- For the INSEE dataset, I've computed the rescaled growth rates $\tilde{g}$ at for $\Delta t = 1$ year and for the underlying variable x as the firm's net sales only.
- For the CRSP/Compustat dataset, thanks to the longer timespan and the quarterly sampling frequency, I have been able to compute the rescaled growth rates of the firms' net sales, total equity, total assets and and market capitalisation (in which case the rescaled growth rates correspond to the normalized returns associated to the firm) and for different sampling frequencies Δt ranging from 1 quarter to 2 years.

In the second case, the reasoning is that if the growth rates have been computed properly, then one should see a gaussian central region that extends as Δt grows larger. As the INSEE dataset consists of a larger number of firms but with a lower sampling frequency, if we see a similar density between the net sales growth rates between the two datasets, then

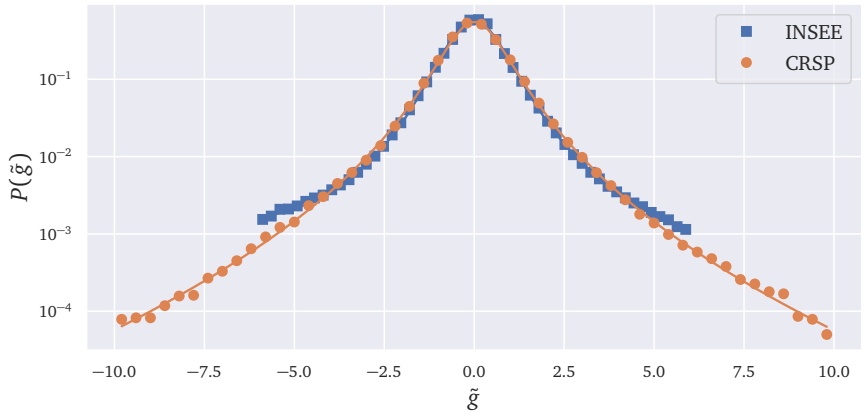


Figure 6.12: Rescaled yearly growth rates $\tilde{g}$ for the net sales in the CRSP/Compustat and INSEE datasets. The circles and squares represent the empirical values for the density, while the solid lines represent the fit with the function defined in Eq. (6.19). Not only does our phenomenological fit work very well, but the two rescaled densities are remarkably similar.

this can also be considered as a justification for our approach, in that it properly characterizes the way firms grow.

From this, one sees that growth rates do indeed present a gaussian central region that grows larger as the period Δt grows larger. There still are stretched exponential tails, but it is not clear whether they are present because of a misestimation of the volatility used in the denominator in Eq. (6.17) or because of a more deep reason. Following Ref. [55], I introduce phenomenologically a density function that is inspired on the Subbotin family, albeit with a gaussian instead of a Laplace central region, reading

$$p_{\tilde{g}}(\tilde{g}) = C \exp\left(- \frac{(\tilde{g} - r_0)^2 / 2\sigma_0^2}{1 + (\tilde{g}/b_p)_+^{2-\mu_p} + (\tilde{g}/b_n)_-^{2-\mu_n}} \right), \tag{6.19}$$

where C is a normalization constant and $(x)_\pm = |x| \mathbf{1}(\pm x \geq 0)$.

This function verifies

$$p_{\tilde{g}}(\tilde{g}) \approx \begin{cases} C_0 \exp\left(-\frac{(\tilde{g}-r_0)^2}{2\sigma_0^2} \right) & \text{for} \quad |\tilde{g}| \ll b_p, b_n \\ C_p \exp\left(-(b_p \tilde{g})^{-\mu_p} / (2\sigma_0^2) \right) & \text{for} \quad \tilde{g} \gg b_p \\ C_n \exp\left(-(b_n \tilde{g})^{-\mu_n} / (2\sigma_0^2) \right) & \text{for} \quad \tilde{g} \ll -b_n \end{cases}, \tag{6.20}$$

so that one sees that this density has a gaussian central region with stretched exponential tails, while allowing an exponent μ_p in the right tail and an exponent μ_n in the left tail. The parameters b_n and b_p therefore control the extent of the gaussian region for negative and positive growth rates: the larger they are, the more gaussian the distribution.

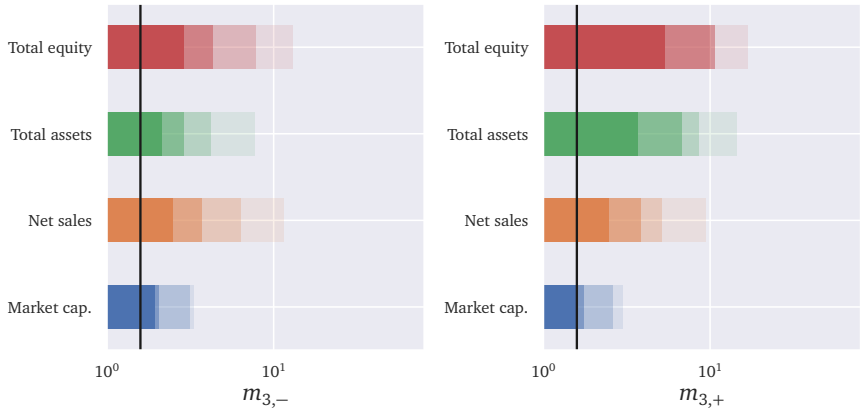


Figure 6.13: Plot of the different moments $m_{3,\pm}$ for different underlying variables x. The increasing shade corresponds to $\Delta t = 1, 2, 4$ and 8 quarters. The solid black line corresponds to the gaussian limit. One sees that the right tail of the market capitalization converges very quickly to a gaussian, while the left tail does so more slowly. In all, all variables show a convergence to a gaussian distribution, even if it happens asymmetrically and at varying speeds for each variable.

I show the results of such a fit for the cases described above on Figure 6.12. The fit with the phenomenological density proposed in Eq. (6.19) is remarkable, and better than that one would obtain with e. g. the Student distribution, although I should like to stress that the tail parameters μ_p and μ_n, while often close to $1/2$, fluctuate a lot when considering different quantities, such as the total equity or total assets of a firm. In the case of the market capitalization no fit with Eq. (6.19) is possible, mainly because the gaussian region seems to have engulfed the whole distribution at the scale of one quarter, and other distributions should be tried instead.

It is however possible to inspect the properties of the distributions of $\tilde{g}$, and in particular check if there is convergence to a gaussian distribution when looking at larger values for Δt. A very simple way to do this, without resorting to parametric fits, is to compute the third moment of the distribution conditioned on the sign, that is:

$$m_{3,\pm}(x, \Delta t) = \mathbb{E}\left(\tilde{g}^3_{x,\Delta t} \,\middle|\, \pm\tilde{g}_{x,\Delta t} > 0\right), \tag{6.21}$$

for different variables x and lags Δt, and comparing with the gaussian result $m_{3,\pm} = \sqrt{\frac{2}{\pi}} \int_0^\infty \mathrm{d}x\; x^3 e^{-x^2/2} = \sqrt{\frac{8}{\pi}}$.

In fact one can also check that the function in Eq. (6.19) captures this as well, by plotting the values of b_p and b_n for different lags Δt and for the variables under consideration in the CRSP/Compustat dataset. I show the results in Figure 6.14, highlighting that the gaussian region does get larger and larger as Δt grows for the variables under consideration.

As said before, the reason behind this convergence is that, for example, the growth rate at one year is statistically equivalent to the sum of the four growth rates corresponding to each quarter in that year. Note also that

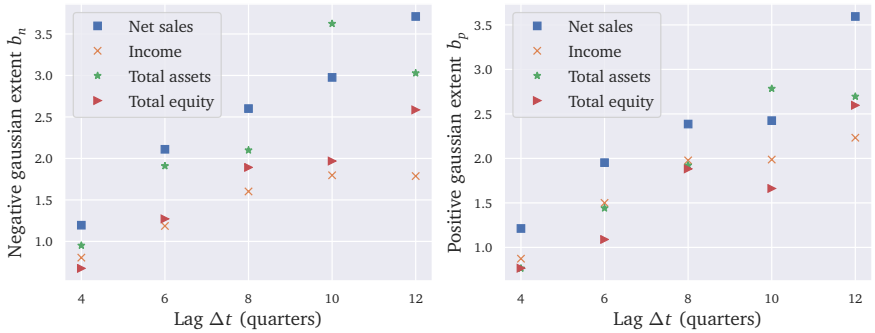


Figure 6.14: Values of b_p and b_n for different variables in the CRSP/Compustat dataset and for growing lags Δt. One sees that both b_p and b_n grow as Δt grows, signalling that the gaussian region of these growth rates becomes larger and larger.

$\tilde{g}$'s units are in terms of the standard deviation of one firm. The results in Figure 6.14 show that the gaussian approximation is already valid at the scale of about two standard deviations for nearly all the variables.

The fact that we see different convergence speeds to a gaussian distribution for the growth rates motivates doing a statistical analysis of the dynamical features of these growth rates.

6.5 THE DYNAMICS OF GROWTH RATES

The main idea behind the introduction of the $\tilde{g}$ variables is that by removing the influence of a firm's idiosyncratic volatilities it is possible to obtain quantities that are comparable between one another. The variable $\tilde{g}^i_{x,\Delta t}(t)$ therefore contains information on the fluctuations that affected firm i at time t, after discarding the information on the typical magnitude of these fluctuations, given by $\sigma^i_{x,\Delta t}$.

It is then possible to study the relationship between, say, its value at a time $t+\tau$ and its value at t by computing correlations: usually, this would be a relation that would be very hard to elucidate given the short time series we have on firm-level data, but the fact that we have removed the heterogeneity between firms means that we could first compute these correlations *for each firm* and then later *average them* across firms to understand the generic behaviour of these growth reads. This contrasts with what one can usually do with financial data, which consists in running very long time averages for a single asset.

I fix henceforth the scale Δt to 1 quarter to use the highest available sampling frequency, and restrict this study to the CRSP/Compustat dataset *only*. The lag τ will therefore be implicitly always in quarters.

For a firm i, I then define the lagged correlator linking quantities the growth rates of quantities x and y as

$$C^i_{x,y}(\tau) = \mathbb{E}_t\left[\tilde{g}^i_x(t+\tau)g_y(t)\right], \tag{6.22}$$

that is the covariance between $\tilde{g}_x^i$ and $\tilde{g}_y^i$[7]. I thus obtain one correlation function for each firm and each couple of variables (x, y). The next step is to average these correlation functions across all firms to obtain a correlation function describing the coupled behaviour of variables (x, y):

$$C_{x,y}(\tau) = \mathbb{E}_i\Big[C^i_{x,y}(\tau)\Big], \tag{6.23}$$

this function, and in particular the autocorrelation functions $C_{x,x}(\tau)$, encode the dynamics of the growth rates.

Another way to look at these dynamics is through the *volatility signature* of these quantities. To understand what it is, I invite the reader to consider again a thought experiment: imagine having a very long time series $x^i(t)$ describing the evolution of a quantity x for firm i. Using the definition of the growth rate g, one could always write, using an initial time $t = 0$ and a lag $\Delta t = 1$,

$$\log\big(x^i(t)\big) = \log\big(x^i(0)\big) + \sum_{k=0}^{t-1} g^i_x(k), \tag{6.24}$$

and compute an object that is called the variogram $\mathscr{V}$ of the time-series $\{\log\big(x^i(t)\big)\}$, namely

$$\mathscr{V}^i(\tau) = \mathbb{E}_t\Big[\log\big(x^i(t+\tau)/x(t)\big)^2\Big] - \Big(\mathbb{E}_t\Big[\log\big(x^i(t+\tau)/x(t)\big)\Big]\Big)^2. \tag{6.25}$$

This function can in turn be rewritten as

$$\mathscr{V}^i_x(\tau) = \big(\sigma^i_x\big)^2\left(\tau + \sum_{k=1}^{\tau}(\tau - k)\, C_{x,x}(k)\right), \tag{6.26}$$

with $C_{x,x}$ the autocorrelation function of the growth rates (assuming here for simplicity that when we compute $\tilde{g}^i_x$ we obtain a time series of unit variance). If the growth rates are uncorrelated then one obtains a linear variogram, while the presence of correlations modify this. One instead plots what is called a volatility signature plot, defined here as

$$v^i_x(\tau) = \sqrt{\frac{\mathscr{V}^i_x(\tau)}{\big(\sigma^i_x\big)^2\,\tau}} = \sqrt{1 + \sum_{k=1}^{\tau}\left(1 - \frac{k}{\tau}\right) C_{x,x}(k)}, \tag{6.27}$$

which has the advantage of giving a clear picture of the dynamics of the underlying time series. Indeed, if $v^i_x(\tau)$ increases as τ increases it is a sign of *trend-following*, meaning that if x grows at time t it is likely to keep on growing, while if it decreases it is a sign of *mean-reversion*.

7 I remind that both series have, by construction, zero mean. Their variance is however not necessarily equal to 1, as the "leave one out" rescaling is not equivalent to dividing by the standard deviation of the time series if there are extreme events.

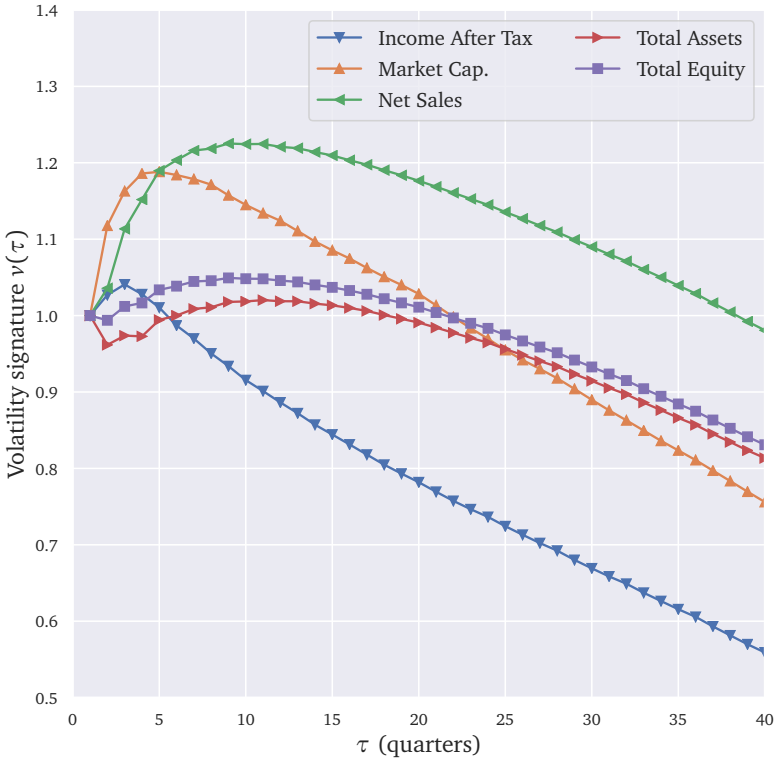


Figure 6.15: Volatility signature plots for various quantities x computed in the CRSP/Compustat dataset. The curves increase when there is trend-following, and decrease when there is mean reversion.

In practice, I compute the volatility signature after computing the correlation functions $C^i_{x,x}(\tau)$ using the last expression on the right hand side of Eq. (6.27). I therefore obtain a function $\nu^i_x(\tau)$ for each firm, and again the fact that I have removed the heterogeneity from the σ^i_x volatilities means that I can aggregate them into a single function describing variable x, as

$$\nu_x(\tau) = \mathbb{E}_i\left[\nu^i_x(\tau)\right]. \tag{6.28}$$

I have computed these volatility signatures and plotted them on Figure 6.15. A striking result is the different behaviour of the market capitalization and other fundamental variables. For the market capitalization, there is trend following up to about 1 year, when mean reversion kicks in and moderates the exuberance of the trend.

This feature was highlighted recently in Ref. [60], with [86, 174] providing possible explanatory mechanisms. Essentially, agents on financial markets overreact on short time-scales (of the order of a few months) as they use simple trend-following strategies that can drive prices away from their "fundamental" value; when the distortion is too large "fundamental-

ists" may kick in and drive the price back to its "true" value, thereby having a moderating, mean-reverting effect on it.

A similar feature appears on firms' fundamental variables, with for example the net sales of a firm first trending and then mean reverting, but the time-scale needed for this change is about the double of that for the market capitalization, and the mean reversion effect is also not as strong. In contrast, the total assets and total equity have more diffusive dynamics, before mean reverting on long time scales, while the income after tax seems to mean-revert strongly from the very beginning, as one would expect from intuition.

This is of course something that can be recovered from the correlation functions. On Figure 6.16 I show the various correlation functions $C_{x,y}$, including the autocorrelation functions $C_{x,x}$. For example, the market capitalisation presents some degree of positive autocorrelation for small values of τ, before presenting a slightly negative autocorrelation indicating mean-reversion.

More interesting is the interplay between the different variables. One sees, for example, that a growth in sales today seems to correlated with a growth in market capitalisation about 4 quarters in the past: one could interpret this in saying that the markets have access to the information on the future sales of a company, through the public knowledge of contracts or such, and react accordingly by setting the price *before* the actual sale is made or reported. Another interesting feature is that a growth in market capitalisation today correlates more strongly with an increase in income one quarter in the future, indicating that an increasing stock price can translate into income for the firm.

6.6 CONCLUSION

The work presented in this section is still ongoing, and constitutes the heart of the empirical work I've done during my thesis. The most important contribution is, I think, the taking into account of heterogeneous volatilities for firm growth. This can be interpreted as a new version of Gibrat's law, in a sense weaker than that given in Eq. (6.3), by claiming that growth rates can be written as

$$g^i_{x,\Delta t} = m_i + \sigma^i_{x,\Delta t}\tilde{g}, \tag{6.29}$$

where the average growth rate m_i and its volatility σ depend entirely on the firm and the quantity under consideration, but where the variable $\tilde{g}$ is roughly considered to be drawn from the same distribution for all firms.

This allows to imagine further work along these lines. Interesting questions are, for example, how the parameters labelled m and σ depend on firm characteristics, like the sectoral dependence shown in Figure 6.7. Another interesting clue is to extend the intra-firm correlations analysis done in Sec. 6.5 to inter-firm correlations, and to try to expose how supply-chain

dependencies translate into sales and production fluctuations, along the lines of the project developed in Chapter 5. The datasets I've worked with in this past Section allow for the exploration of these issues, and I am persuaded that the analysis developed here will bear more fruits.

This concludes the work I did to model and study firms and their growth, although it led me to ask other questions related to economic issues. These are, to name a few, the issue of the existence of an economic equilibrium raised in Chapter 5 and how one may converge to it; the question of how interactions microscopic, agent-level interactions can lead to completely unpredictable collective phenomena as developed in the study of avalanches in Sec. 3.4.2, and finally the impact of heterogeneity and the possibility of having endogenous fluctuations, as raised more broadly in this and the past Chapter.

The following Section is an attempt to study these issues. After what could be described as a macro-economically motivated study on firms, it may seem somewhat of a leap to consider things such as herding behaviour, the exploitation of fishing areas and memory effects in decision theory, but I firmly believe it is necessary to tackle these problems now to have a better and more pertinent outlook towards modelling economic and social phenomena.

At the beginning of Chapter 4, I cited a number of economists who felt that the theoretical arsenal they had at the time of the 2007−2008 crisis was not adapted to handle such a difficult situation. This, in turn, has led to calls to aggressively question a set of concepts that are axiomatic in mainstream economic theory [62], such as the rational representative agent paradigm, ergodicity or the very notion of equilibrium. This suggest looking beyond the infinitely wise and rational *homo economicus* towards more realistic models for *homo sapiens*. This is, in a nutshell, the programme set out by complexity economics [40] that motivated the following Chapters.

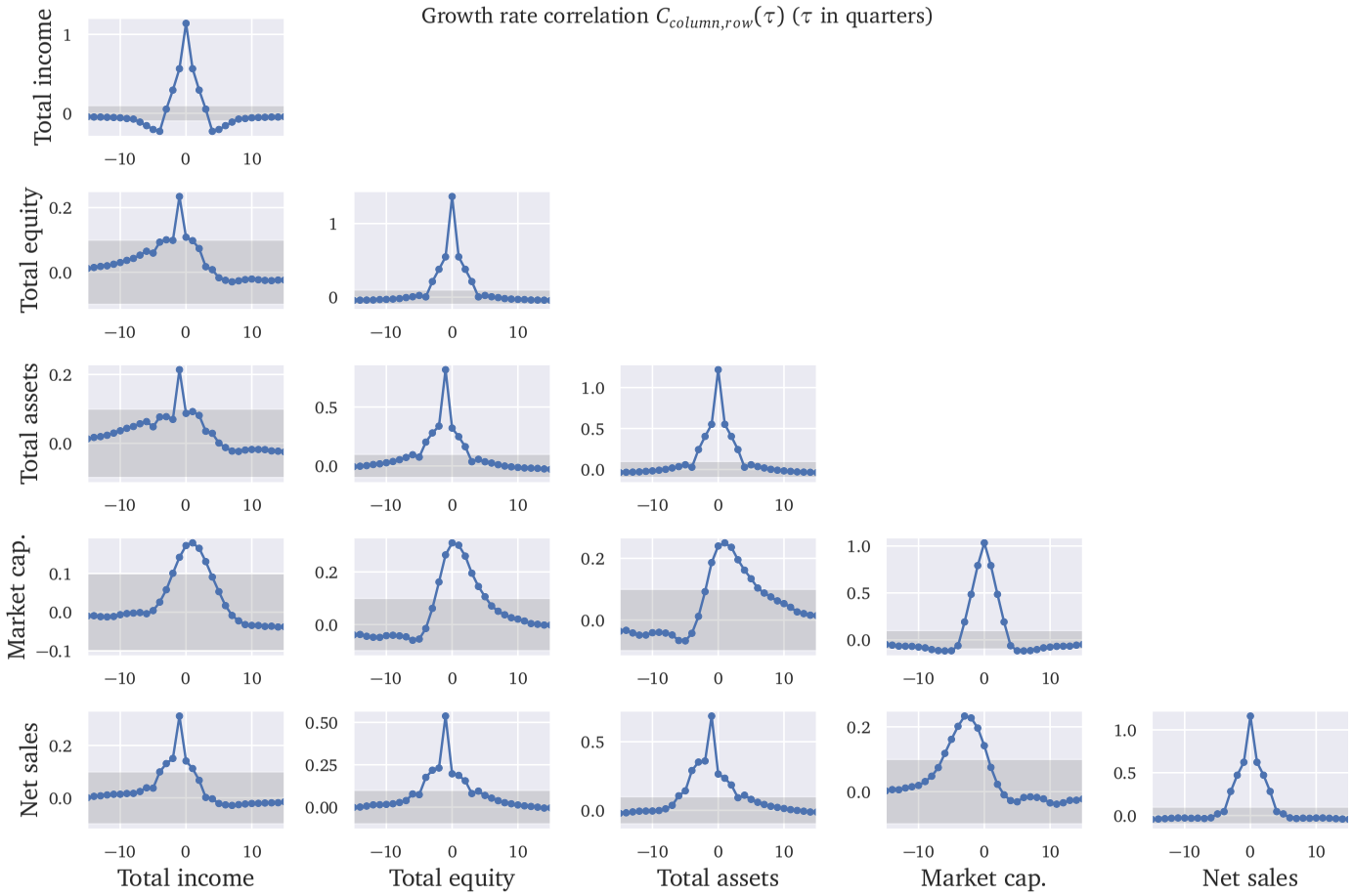


Figure 6.16: Correlation functions $C_{x,y}(\tau)$ for the different variables, computed using the CRSP/Compustat dataset. The shaded grey area corresponds to $(-0.1, 0.1)$.

Part IV

ERGODICITY, MEMORY, IMITATION AND COMPLEX SYSTEMS

7

HERDING

In modelling the aggregate behaviour of agents, traditional economics has often made use of a "representative agent" maximizing an utility function. With this view, many models become analytically tractable and one provides them with what is known as "micro-foundations", although this paradigm has come under strong criticism in the past years. It would also seem to be linked to the idea that if individuals are left to their own devices then the economy will self-organise into a state with satisfactory characteristics. This, according to Kirman, has become an assumption that is not backed by empirical evidence nor theory [158].

Indeed, modern economies consist of millions of agents, alone or organized in groups or firms, that interact together via a market. Representing their aggregate by a single entity throws the baby with the bathwater, as it completely neglects how these interactions may happen, and indeed how coordination or miscoordination may emerge. The temptation to declare the representative agent defunct is therefore quite strong [154].

Already in Sec. 3.4.2 the reader became familiar with the Random Field Ising Model (RFIM): in that model, agents can make a binary choice ± 1, or A and B, but they are influenced by their peers and by an "external field" that captures e. g. all publicly available information. Variations of this external field can then provoke large scale *opinion avalanches*, where one agent changes her mind and ends up influencing others to do the same. Crucially, these spectacular collective rearrangements, similar to *phase transitions* in physics such as the liquid/solid transition for water, only happen because of interactions. This is already a big conceptual leap, that is well summarized in Anderson's anti-reductionist manifesto, *More is different* [17].

Another deeply rooted idea, tight-knit with that of the representative agent, is that of *economic equilibrium*. This idea has, in the words of Farmer and Geanakoplos [109] "an enormous impact on the way economists think, and indeed in many ways defines the way economists think". Within that framework, and as highlighted in Chapter 4, fluctuations about equilibrium values are of *exogenous* nature, meaning that they come from an external "hand of Jupiter", to quote Adam Smith. It would instead seem more natural to take into account mechanisms for *endogenous fluctuations* [63, 245].

In the following, I will present Kirman and Föllmer's[1] *ant recruitment model* [157]. This model combines basic mechanisms of noise and self-

1 Although Föllmer does not appear in the list of authors, it is he who provided the mathematical solution to the model that Kirman came up with, although from Kirman's account Föllmer did not initially deem his contribution to the model mathematically relevant and therefore didn't sign it. At Kirman's request following the publication of Ref. [189], I attribute the model to both of them.

reinforcement, in a spirit very similar to Pólya urns, to create large purely endogenous opinion avalanches.

My contribution to this model and this literature lies in providing, along with Fosset, Benzaquen and Bouchaud, a full dynamical solution to the model, thereby giving some insight into the way in which the above cited avalanches take place, and was published in Ref. [189]. Further work done with Fosset, Benzaquen and Kirman draws heavily from this model to explain empirically observed fishing vessel dynamics in the Adriatic sea. This last study has not yet been published, but the main results are also summarized below.

7.1 THE ANT RECRUITMENT MODEL

The *ant model* [157] undoubtedly stands among the most inspiring toy models in the behavioural economics literature. While initially inspired by the experiment described below, its conclusions have implications much beyond collective animal behaviour, as it has been used to model shifts in sentiment of economic agents, trend reversal in financial markets, herding and social influence, etc. It is also akin to another famous model in population dynamics with competing species: the Moran model [193].

Initially, entomologists were puzzled by the behaviour of ants who, able to choose between two identical and inexhaustible food sources A and B, tended to concentrate on one of them for a while, but occasionally switched collectively to the other without any apparent reason[98, 39]. Such intermittent herding behaviour is very similar to that observed in humans choosing between equivalent restaurants [38], or in financial markets [230, 239, 171], and is consistent with the notion that most large fluctuations are of purely endogenous nature: in contrast with the views of standard economic choice theory, the asymmetric exploitation observed in ants does not seem to correspond to the equilibrium state of an isolated representative ant with rational expectations.

The entomologists' observations suggest instead that such phenomena should rather be explained in terms of interactions between individual agents, ants in this particular setting, through what biologists call recruitment dynamics. Note however that this is also in contrast with the RFIM model presented in Sec. 3.4.2, where avalanches appeared *exogenously* through the modification of the external field. To account for such intricate behaviour, Kirman proposed a simple yet insightful model [157] based on tandem recruitment that I describe below.

Consider the N ants from the experiment and denote by $k(t) \in [0, N]$ the number of ants feeding on source A at time t. The model states that when two ants meet, one of them converts the other with probability μ/N, but each ant may in addition change its own mind spontaneously with probability ε. Within such a simple setting, Kirman was able to show that, in the large N limit, the stationary state depends only on a parameter $\alpha := \varepsilon/\mu$. When $\alpha > 1$ the stationary distribution of k is unimodal, with

a maximum at $k = N/2$, whereas for $\alpha < 1$ it is bimodal, with the most probable states being $k = 0$ and $k = N$ (corresponding to the situation observed experimentally). Remarkably, the interesting $\alpha < 1$ regime can be obtained even for weakly persuasive agents (small μ), provided the self-conversion rate ε is also low.

The most important point to be raised is that in the $\alpha < 1$ regime none of the k states is, alone, an equilibrium state. Although the system can spend a long time at $k = 0, N$ (local stationarity), these states cannot be considered as equilibria: every possible state is always revisited, and there is no convergence to any one of them in particular, discarding also the possibility of having multiple equilibria. Rather, there is perpetual change, and the system's natural endogenous dynamic is only in a *statistical equilibrium*. Contrary to this, as was stated in the introduction of this Chapter, most economic models focus on finding the equilibrium state to which the system under study will finally converge, and say that the latter may only be knocked off its path by large exogenous shocks. As stated also in Chapter 4, this paradigm has come under heavy criticism for its inability to guide policy makers out of "dark corners" [49].

Indeed financial markets, and even larger economic and social systems, display a number of regular large switches (correlations, mood of investors etc.) which do not always seem to be driven by exogenous shocks. In Kirman's stylised setting such switches can be understood endogenously, as in similar models where agents are allowed to interact and influence their peers' behaviour, thereby causing the emergence of non-trivial auto-reinforcing social dynamics [271, 3]. Several extensions of Kirman's model have been proposed [171, 159, 130], and the basic dynamics of imitation coupled with random switching can also be used to describe agents making political choices – say voting to the left or to the right – while being influenced by their peers, as described in the so-called voter model, see e.g. [163].

The sections below present a continuous description of the ant model which notably allows to derive the typical switching time, using classical methods from statistical physics and quantum mechanics.

7.1.1 *Master Equation*

As mentioned above, the original model describes N ants faced with two identical food sources, with the relevant dynamical variable being k, the number of ants feeding on – say – source A. Each time step allows an ant to either switch randomly to the other food source with probability proportional to ε, or to get recruited by another ant from the other food source with probability proportional to μ.

Defining the unit of time as the time required for all the ants to make a decision leads to $\mathrm{d}t = 1/N$ as the infinitesimal time unit. It is also clear that, to remain intensive in the large N limit, the probability to interact with another ant should be proportional to $1/N$. Altogether, the Master

equation for the evolution of the probability $\mathbf{P}(k,t)$ that there are k ants feeding at source A at time t reads:

$$\begin{aligned}\mathbf{P}\left(k,t+\frac{1}{N}\right)-\mathbf{P}(k,t) = \frac{1}{N}\Big\{&W(k+1\to k)\mathbf{P}(k+1,t)\\ &+W(k-1\to k)P(k-1,t)\\ &-\big[W(k\to k-1)-W(k\to k+1)\big]\mathbf{P}(k,t)\Big\},\end{aligned} \tag{7.1}$$

where the transition rates are given by:

$$\begin{aligned}W(k\to k+1) &= \left(1-\frac{k}{N}\right)\left(\varepsilon+\frac{\mu}{N}\frac{k}{N-1}\right)\\ W(k\to k-1) &= \frac{k}{N}\left(\varepsilon+\frac{\mu}{N}\frac{N-k}{N-1}\right).\end{aligned} \tag{7.2}$$

Note that this specification only differs from Kirman's original one in the rescaling of the recruitment rate by N. With the notations of [157], $1-\delta = \mu/N$. Therefore, μ is the average fraction of ants deciding to imitate their peers per unit time.

7.1.2 *Continuous description and Fokker-Planck equation*

This section follows Kirman's original paper [157], and proposes a way to derive a proper continuous-time Fokker-Planck equation in the limit $N\to\infty$. Define the variable $x=\frac{k}{N}\in[0;1]$ together with its probability density function $f(x,t)$.

Taking the continuous limit $N\to\infty$ of Eq. (7.1), which amounts to taking the continuous limit of the transition matrix as I've explained in Sec. B.2 and that is explicitly done in Appendix E.1, leads to the following Fokker-Planck equation:

$$\partial_t f = \partial_x J^f, \quad \text{with} \quad J^f(x,t) = -\varepsilon(1-2x)f(x,t)+\mu\partial_x\left[x(1-x)f(x,t)\right], \tag{7.3}$$

the probability flux. The conservation of the number of ants in the model is ensured by the condition $J^f(x,t)=0$ at the boundaries $x=0$ and $x=1$ at all times.

Equation (7.3) corresponds to the following stochastic process for x:

$$\dot{x} = \varepsilon(1-2x)+\sqrt{2\mu x(1-x)}\eta(t), \tag{7.4}$$

with η a Gaussian white noise with unit variance. One can note that while the drift term $\varepsilon(1-2x)$ is maximal at the boundaries and tends to pull x towards $1/2$, the noise term has the opposite effect. The diffusion constant is proportional to $\sqrt{2\mu x(1-x)}$ and is maximal at $x=1/2$ and so tends to push the system away from $x=1/2$.

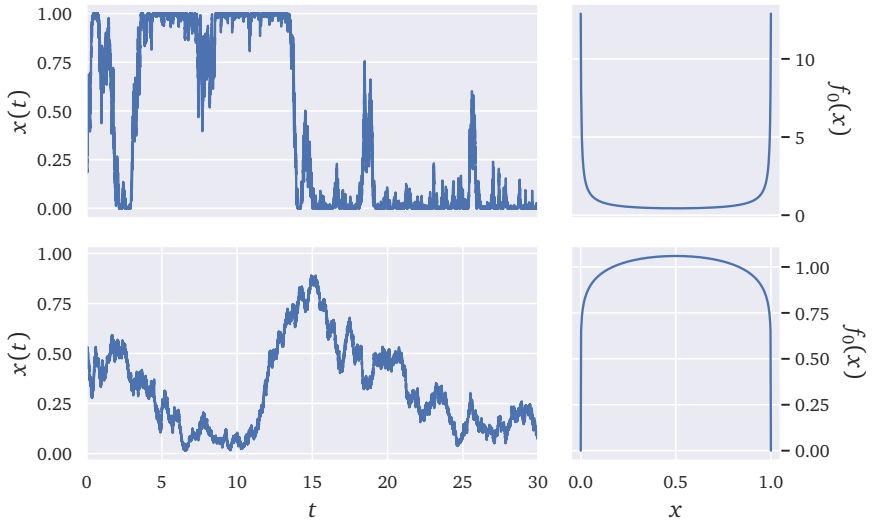


Figure 7.1: Simulations of the model in the continuous limit. The top plots correspond to $\alpha = 0.1 < 1$ while the bottom ones to $\alpha = 2 > 1$. Both simulations were run with $\varepsilon = 0.1$. The left panels display the evolution of $x(t)$ as defined in Eq. (7.4). The right panels display the corresponding stationary probability densities, as given by Eq. (7.5).

Note that this stochastic process is very similar to the Moran model of genetic population dynamics [193] — with the same diffusion term $\propto \sqrt{x(1-x)}$ — where the analogue of the number of ants at each food source is the proportion of genes from two competing alleles (A or B) [278]. The ε term corresponds to spontaneous mutations. When $\varepsilon = 0$, there is a non zero probability that the whole population becomes of type A or B after a finite time, corresponding to $\delta(x)$ or $\delta(1-x)$ contributions to $f(x,t)$ with a time dependent weight. Refs. [153], and [180] give a recent thorough discussion of this situation.

Interestingly, one can associate different pay-offs to picking A or B, leading to a game-theoretical extension of this family of models, see [19] for a development in the context of evolutionary games. It is also equivalent to a model describing noise-induced bistability transitions in chemical reactions [111], as shown in [91], where the equation describing the process is equivalent to Eq. (7.3). Finally, the same diffusion term $\propto \sqrt{x(1-x)}$ also emerges naturally within the so-called voter model, that describes the opinion dynamics of peer-influenced voters [163].

When $\varepsilon > 0$, one can check that the normalised stationary distribution $f_0(x)$, obtained by setting $J^f(x,t) = 0$, writes:

$$f_0(x) = \frac{\Gamma(2\alpha)}{\Gamma^2(\alpha)}\big[x(1-x)\big]^{\alpha-1}, \quad \text{with} \quad \alpha := \frac{\varepsilon}{\mu}. \tag{7.5}$$

This result is the same as that obtained by Föllmer and Kirman in [157].

Upon looking at the behaviour of the solution, shown in Fig. 7.1, one can see that there is a clear transition in the behaviour of the model at $\alpha_c = 1$.

For $\alpha > \alpha_c$, the stationary density in Eq. (7.5) is maximal at $x = 1/2$, and the dynamics show that $x(t)$ fluctuates around $1/2$, corresponding to a situation where the ants are, on average, evenly distributed across both food sources. For $\alpha < \alpha_c$ the density f_0 diverges at the boundaries.

The top left panel in Fig. 7.1 shows that this corresponds to a very different picture, in which nearly all of the ants choose either one of the sources for a certain amount of time, until a noise-induced "avalanche" causes a switch over to the other source. It is also easy to check that in the absence of noise (and $\alpha \to 0$) the long-time stationary density is given by $f_0(x) = p\delta(x) + (1-p)\delta(x-1)$, a situation discussed at length in [180].

Having this in mind, a natural question to ask is: Given a certain initial condition $f(x,0) = \delta(x - x_0)$, how long does it take for the system to converge to the stationary state, or equivalently, how long does it take for the ants to switch from one source to the other in the $\alpha < 1$ regime? The answer to this question is my contribution to the study of this and other related models.

7.1.3 *Schrödinger's equation and general solution*

Here, a full dynamical solution in terms of the eigenvalues and eigenfunctions is obtained in terms of the eigenvalues and eigenfunctions of a certain quantum mechanical Hamiltonian.

Using the Itô rule (see Ref. [145] and Appendix B), one can see that introducing a change of variables $\varphi(x)$ in Eq. (7.4) yields a noise term proportional to $\sqrt{x(1-x)}\varphi'(x)$, and so motivates a choice satisfying $\varphi'(x) = 1/\sqrt{x(1-x)}$. This led me to define a new, more convenient, variable $\varphi \in [-\pi/2, \pi/2]$ as:

$$\sin\varphi = 2x - 1. \tag{7.6}$$

The corresponding Fokker-Planck equation for its probability density $g(\varphi, t)$ writes:

$$\begin{aligned} \partial_t g = \mu \partial_\varphi J^g, \quad \text{with} \quad & J^g(\varphi,t) = 2\beta \tan\varphi\, g(\varphi,t) + \partial_\varphi g(\varphi,t), \\ \text{and} \quad & \beta := \alpha - \frac{1}{2}, \end{aligned} \tag{7.7}$$

where the probability flux must now verify $J^g(\pm\pi/2, t) = 0$ at all times. Note however that this is only exact in the limit $\tan(\varphi) \ll \frac{N}{2\alpha}$, as discussed in Appendix E.2. Setting again $J^g = 0$ everywhere, one finds the normalised stationary solution:

$$g_0(\varphi) = \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{\sqrt{\pi}\Gamma(\alpha)} (\cos\varphi)^{2\alpha - 1}. \tag{7.8}$$

The advantage of this formulation in φ is that, in contrast with the former, the second order derivative term $\partial_{\varphi\varphi}$ in Eq. (7.7) only depends on φ through $g(\varphi, t)$, as diffusion is no longer position-dependent. Standard

techniques for the resolution of Fokker-Plank equations, see e. g. [226] and Appendix B.2.2, motivate the introduction of a function Ψ such that:

$$g(\varphi,t) := \sqrt{g_0(\varphi)}\Psi(\varphi,t)\,, \tag{7.9}$$

and $\Psi(\varphi,t) \to \sqrt{g_0(\varphi)}$ when $t \to \infty$.

Combining Eqs. (7.7) and (7.9) one obtains a Schrödinger-like equation of the form [89]:

$$-\frac{1}{\mu}\partial_t \Psi = \mathbf{H}\Psi, \tag{7.10}$$

where the Hamiltonian $\mathbf{H}$ is defined as:

$$\mathbf{H} := -\partial_{\varphi\varphi} + V(\varphi), \quad V(\varphi) := -\beta + \beta(\beta-1)\tan^2\varphi, \tag{7.11}$$

and with boundary conditions given by:

$$\left[\cos^\beta\varphi\big(\beta\tan\varphi\Psi(\varphi,t) + \partial_\varphi\Psi(\varphi,t)\big)\right]_{\varphi=\pm\pi/2} = 0\,. \tag{7.12}$$

I have left the μ parameter out of the Hamiltonian $\mathbf{H}$ in order to ease the comparison to the canonical form presented in [201, 261]. The $\tan^2$ term in Eq. (7.10) is known as the Pöschl-Teller potential [201], and appears in a similar context of social dynamics within a version of the voter model in Ref [264]. The potential here was fully solved in the case $\beta > 0$ with boundary conditions $\Psi(\pm\pi/2,t) = 0$ in [261].

To be applicable to the framework under study, I now verify that their solutions also satisfy Eq. (7.12) in the general case $\beta > -1/2$. As discussed in Sec. B.2, the Hamiltonian $\mathbf{H}$ is hermitian (contrarily to the Fokker-Planck operator) and has a discrete set of orthogonal eigenfunctions and eigenvalues, given by:

$$\mathbf{H}\Psi_n = \mathscr{E}_n\Psi_n\,, \tag{7.13}$$

where, splitting into even ($n = 2k$) and odd ($n = 2k+1$) states:

$$\begin{aligned}
\mathscr{E}_n =& n(2\alpha+n-1)\,,\\
\Psi_{2k}(\varphi) =& A_{2k}(\beta)\ {}_2F_1\left(-k,\beta+k;\beta+\tfrac{1}{2},\cos^2\varphi\right)\cos^\beta\varphi\,,\\
\Psi_{2k+1}(\varphi) =& A_{2k+1}(\beta)\ {}_2F_1\left(-k,\beta+k+1;\beta+\tfrac{1}{2},\cos^2\varphi\right)\sin\varphi\ \cos^\beta\varphi\,,
\end{aligned} \tag{7.14}$$

with ${}_2F_1$ the ordinary hypergeometric function.[2] The coefficients A_n are set such as to ensure normalisation, $\int_{[-\pi/2,\pi/2]}\Psi_n\Psi_m = \delta_{n,m}$, and can be expressed as integrals of hypergeometric functions. Note that the parity of

2 Here, the function ${}_2F_1$ takes the form of a polynomial: ${}_2F_1(-k,a;b,u) = \sum_{l=0}^{k}\binom{k}{\ell}(-1)^\ell\frac{\Gamma(a+\ell)}{\Gamma(a)}\frac{\Gamma(b)}{\Gamma(b+\ell)}u^\ell$ for any positive integer k.

n also defines the parity of the function Ψ_n with respect to the y-axis. One can then easily check that for all n (both even and odd):

$$\cos^\beta\varphi\left[\beta\tan\varphi\Psi_n(\varphi)+\Psi_n'(\varphi)\right]\underset{\varphi\to\pm\pi/2}{\approx}\left(\frac{\pi}{2}\mp\varphi\right)^{1+2\beta}, \tag{7.15}$$

which, since $\beta > -1/2$, ensure that the boundary conditions given by Eq. (7.12) are satisfied. Noting that $\Psi_0 = \sqrt{g_0}$, the general solution of Eq. (7.10) then reads:

$$\Psi(\varphi,t) = \lambda_0\sqrt{g_0(\varphi)} + \sum_{n>1}\lambda_n\Psi_n(\varphi)e^{-\mu\mathscr{E}_n t}, \tag{7.16}$$

with λ_n given by the projections of the initial conditions on each mode n, namely $\lambda_n = \int_{-\pi/2}^{\pi/2}\mathrm{d}\varphi\ \Psi_n(\varphi)\Psi(\varphi,0)$.

Back to the physical variable x, the initial condition $f(x,0) = \delta(x - x_0)$ becomes $g(\varphi,0) = \delta(\varphi-\varphi_0)$ with $\varphi_0 = \arcsin(2x_0-1)$. Further using Eq. (7.9), it is easy to see that the initial condition in turn translates into $\Psi(\varphi,0) = \delta(\varphi-\varphi_0)/\sqrt{g_0(\varphi)}$. The full solution for $g(\varphi,t)$ follows:

$$g(\varphi,t) = g_0(\varphi) + \sum_{n>1}e^{-\mu\mathscr{E}_n t}\Psi_n(\varphi_0)\sqrt{g_0(\varphi)}\Psi_n(\varphi), \tag{7.17}$$

with the orthogonality between $\Psi_0 = \sqrt{g_0}$ and Ψ_n for $n > 1$ ensuring that $\int_{-\pi/2}^{\pi/2}\mathrm{d}\varphi\ g(\varphi,t) = \int_{-\pi/2}^{\pi/2}\mathrm{d}\varphi\ g_0(\varphi) = 1$, or equivalently for $f(x,t)$:

$$f(x,t) = f_0(x) + \sum_{n>1}e^{-\mu\mathscr{E}_n t}\Psi_n(\varphi_0)f_n(x), \tag{7.18}$$

with:

$$f_n(x) = \frac{\sqrt{g_0(\varphi(x))}\,\Psi_n(\varphi(x))}{2\sqrt{x(1-x)}}, \tag{7.19}$$

(see Appendix E.3.2 for an explicit expression). Equation (7.18) is the central result of the present communication.

7.1.4 *Relaxation towards the stationary state*

With the full dynamical solution of Eq. (7.18) at hand, one can see how long a system initially prepared at an initial value $x_0 \approx 0$, for example, takes to explore the whole space. In other words, one can ask how much time τ is required to reach, say, $x(\tau) \approx 1$ with a reasonable probability.

Since the stationary distribution f_0 has weight on the whole interval $[0;1]$, this time τ is none other than the relaxation time (or ergodic time) τ_R required to converge to stationarity. Owing to the form of Eq. (7.18) this convergence is asymptotically exponential, with the slowest mode given by $n = 1$. Hence, it is given by:

$$\tau_\mathrm{R} := \frac{1}{\mu\mathscr{E}_1} \equiv \frac{1}{2\varepsilon}. \tag{7.20}$$

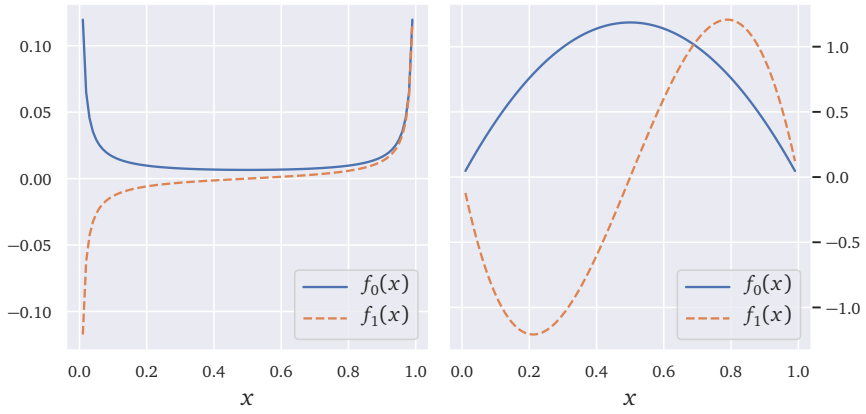


Figure 7.2: A plot showing the shape of the first two modes $f_0(x)$ and $f_1(x)$. The panel on the left corresponds to $\alpha = 0.1$, while the one on the right corresponds to $\alpha = 2$. $f_0(x)$ is the stationary state, whereas $f_1(x)$ is the slowest decaying mode and corresponds to hopping between the two food sources.

Perhaps surprisingly, this relaxation time depends only on the spontaneous switching rate ε, but not on the recruitment intensity μ. Since $n = 1$ corresponds to the slowest mode of the system, it also governs the collective "switch time" between the two food sources, A and B – see Fig. 7.2.

I have checked the prediction for the switching time numerically by running trajectories starting at $x_0 = \Delta x \ll 1$ and computing the probability $\mathbb{P}(x(t) > 1-\Delta x)$. This quantity should converge to $\int_{[1-\Delta x;1]} f_0$ at an exponential rate $\propto e^{-\mu \mathscr{E}_1 t}$, which is in perfect agreement with the simulations, see Figure 7.3.

Similarly, given an initial condition $x_0 = 1/2$ where the ants are initially distributed evenly between the two sources, one may ask how long it takes for all the ants to "decide" on concentrating on one of them. Since this condition is equivalent to $\varphi_0 = 0$, and since Ψ_1 is an odd function of φ, it follows that $\Psi_1(\varphi_0) = 0$ in this case. The convergence to the stationary distribution is then controlled by the second mode, with a much shorter relaxation time given by:

$$\tau'_{\mathrm{R}} := \frac{1}{\mu \mathscr{E}_2} \equiv \frac{1}{4\varepsilon + 2\mu}. \tag{7.21}$$

Directly applying tools from stochastic calculus on Eq. (7.4), one can obtain the following correlation functions (see Appendix E.4):

$$\mathrm{Cov}\left[\sigma_n(x(T+t)), \sigma_n(x(T))\right] \propto e^{-\mu \mathscr{E}_n t}, \tag{7.22}$$

where $\sigma_n(x)$ are polynomials of degree n that allow one to "diagonalize" the evolution of the correlations:

$$\begin{aligned} \sigma_1(x) &= x, \\ \sigma_2(x) &= x(1-x), \\ \sigma_3(x) &= (2x-1)\left[\left(1+\frac{2\alpha}{3}\right)(2x-1)^2 - 1\right]. \end{aligned} \tag{7.23}$$

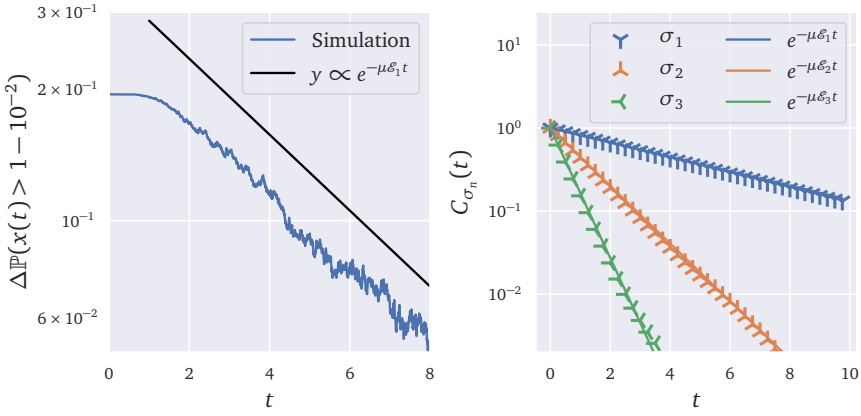


Figure 7.3: Left: plot of $\Delta\mathbb{P}(x(t) > 1-10^{-2})$, defined as the difference between $\mathbb{P}(x(t) > 1-10^{-2})$ and its stationary value, for $\varepsilon = 0.1$ and $\mu = 0.5$. The initial condition is $x_0 = 10^{-2}$. Right: plots of the covariances $C_{\sigma_n}(t) = \mathrm{Cov}\left[\sigma_n(x(T+t)), \sigma_n(x(T))\right]$, computed over simulations with $\varepsilon = 0.1$ and $\mu = 0.2$. The agreement with theoretical predictions is excellent.

See Appendix E.4 for further details and Figure 7.3 for a comparison with numerical results.

This result actually hides a deeper interpretation of the different modes f_n. In the case described above, one can surmise that the dynamics of the moments $\mathbb{E}[x]$, $\mathbb{E}[x^2]$ and $\mathbb{E}[x^3]$ are determined exclusively by the modes f_1, f_2 and f_3. In fact, focusing on any moment $\mathbb{E}[x^m]$, it is possible to prove that:

$$\forall n > m, \quad B_{n,m} = \int_0^1 \mathrm{d}x \; f_n(x) x^m = 0\,, \tag{7.24}$$

as well as for all values n that do not have the same parity as m. This implies in fact that the dynamics of the moments $\mathbb{E}[x^m]$ are fully described by the modes $(f_1, \ldots, f_m)$, with only even values of n contributing to even moments m and vice-versa. For example, for $m = 3$ with the initial condition $x(0) = x_0$ one may compute:

$$\mathbb{E}[x^3(t)] = B_{1,3}\Psi_1(\varphi_0)e^{-2\varepsilon t} + B_{3,3}\Psi_3(\varphi_0)e^{-3(2\varepsilon+2\mu)t}, \tag{7.25}$$

where the exact expression of $B_{n,m}$ is given in Appendix E.3.3, Eqs. (E.34) and (E.35). Mind that $B_{0,m}$ is the stationary value of moment $\mathbb{E}[x^m(t)]$ for all moments.

7.1.5 *Perspectives*

In the last Section I have shown that this very simple model can be solved *entirely* using comparatively simple analytical tools. These techniques, I think, could be of huge benefit in the emerging field of "complexity

economics" as a way of understanding the dynamics of agent-based models. Ref. [121] for example uses similar techniques to study the dynamics of wealth inequality.

The most important insight from this solution is that the switch time between two sources depends only on the "spontaneous conversion" rate ε, and *not* on the recruitment rate μ. This means that a single ant deciding on its own to explore an alternative food source can trigger an "avalanche" where the whole colony follows suit. It is also my hope, and that of my collaborators, that these techniques can be used to solve more general models of Pólya urns.

Another interesting question that emerges when studying this model is how it can be expanded to model how a set of agents – the ants – can exploit sources of perishable resources – the food sources, that would no longer be infinitely plentiful. This is the purpose of the following section, where, by inspiration from illuminating past work on fish markets [155, 125, 156], I studied the dynamics of fishing vessels along with Antoine Fosset, Michael Benzaquen and Alan Kirman.

7.2 A SIMPLE MODEL OF HERDING AND THE EXPLOITATION OF FINITE RESOURCES

The following constitutes work that is yet to be published. It is the study of a problem of general interest is that of the individual and collective exploitation of a resource, by attempting to take into account the aggregate behaviour of individuals and the type of herd behaviour I have described in the past section. In a realistic setting, contrary to what happens in the original ant experiment where the food sources cannot be depleted, individuals may herd and exploit collectively a given resource without taking into account the overall consequence of their actions. This can, for example, lead to complete exhaustion of said resource in, say, an agricultural setting. It's what has been called "The Tragedy of the Commons" [139].

Here, I focus in the case of fisheries, with vessels exploiting a set of *fishing areas*. Past literature has tried to understand the strategies used by individual boats to decide when and where to fish. A simple idea is, for example, that boats fish until their catch falls below a certain threshold and then move on [127]. This is an individualistic behaviour that is in contrast observed in Ref. [266], where they claim that "there is evidence that vessels make decisions about where to fish based on both their own recent catch history and on observation about the location and aggregation of other vessels", all the while claiming that collective behaviour is not the main drive of the observed dynamics.

In another take, Allen and McGlade [6] developed models based in part on Lotka-Volterra dynamics, similar to those given in Chapter 5 and in particular in Eq. (5.12). They emphasised "the importance of information exchange in defining the attractivity of a particular fishing zone to different fleets", and highlighted that "models based on global principles, such as

'optimal efficiency' or 'maximum profit', are clearly of dubious relevance to the real world. ", motivating approaches such as the one described below.

The model described here builds on the ant model described in the previous section, where the actors follow simple rules but their interaction can produce interesting dynamics. In the following, I will first present stylized facts from empirical data related to boats from the Italian ports of Ancona and Pescara before presenting the model.

7.2.1 *Description of the data*

I use the Fishing Vessels Dataset from Global Fishing Watch [133] from October 2012 to December 2016. Since the aim is to analyse the behaviour of fishermen seeking to exploit clearly distinguishable fishing areas, the project focused geographically on the Adriatic Sea and specifically on the area encompassing the Italian cities of Ancona and Pescara in which two of the largest fishing harbours and fish markets are set (see for example [125] for a detailed study and description of the Ancona fishing market).

The two cities are separated by a reasonable distance of about 150 km, meaning that boats based in one city can easily find themselves fishing close to the other. Further, the existence of large and comparable fish markets in both cities hints the possibility of matching fishing activity to market data, provided of course one has access to the latter. Note that while another city, San Benedetto del Tronto, lies between Ancona and Pescara, it is responsible for a rather negligible amount of the activity in the area.

The analysis has also been restricted to the behaviour of trawlers. These boats have a low cruise speed and fish in shallow waters close to the coast. A reasonable hypothesis, which was confirmed with the local market authorities, is that trawlers fishing in the area are based in either one of the two cities and go out for a short amount of time before coming back to sell their catch on the local market. In particular my collaborators and I were told that, due to the policy of the market to sell fresh local fish, vessels (almost) always get back to the port after 24 hours. We were also told that while there is no ban for a boat registered in a given port to land his fish elsewhere, this seldom happens.[3] In other words, one expects trawlers based in, say, Pescara to leave port, fish for at most a day or two and then come back to sell their catch.

The reduced data set consists of daily tracking of these trawlers, identified by their 9-digit Maritime Mobile Service Identity (MMSI) number. Each vessel is tracked on a latitude-longitude grid with resolution 0.1.1 squared degrees. At Ancona and Pescara's latitude ($\approx 43°$ North), this implies a spatial resolution of $\approx 11 \times 8$ km^2 (latitude by longitude). Finally, a

3 According to the director of the Ancona fish market, there are no relationships with nearby wholesale markets (Pescara and San Benedetto del Tronto), and, two or three times a year, it happens that a boat based in the nearby port in the north (Fano or Cattolica) comes to sell.

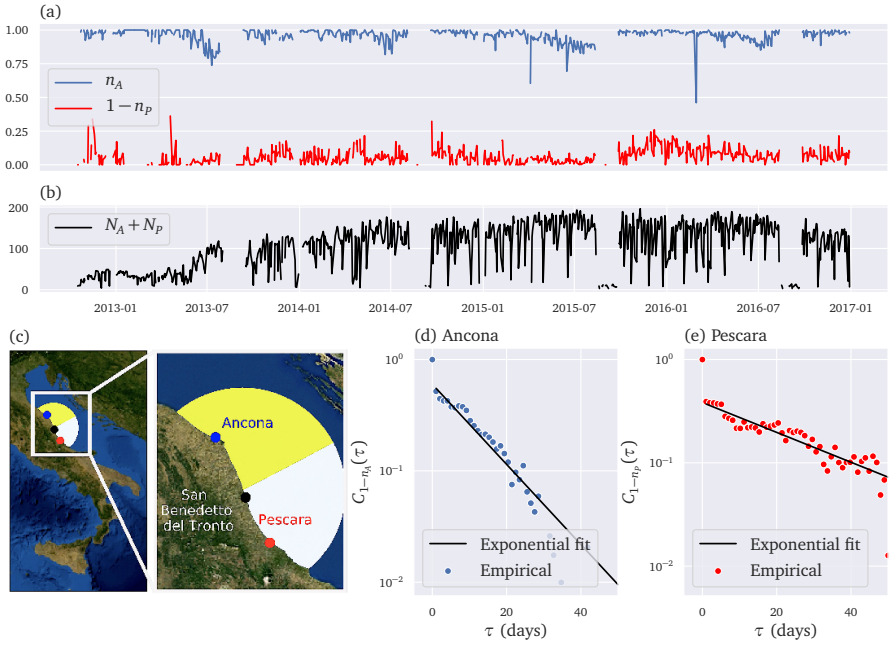


Figure 7.4: Various Figures depicting the data. Blue curves and markers correspond to data related to the area of Ancona, while red curves and markers correspond to Pescara. (a) Plot of the fraction $n_i(t)$ as defined in Eq. (7.27) (b) Plot of the total number of active boats through time $N_A + N_P$. (c) Satellite view of the Adriatic Sea along with the areas $\mathscr{D}_\mathrm{A}$ and $\mathscr{D}_\mathrm{P}$ defined in Eq. (7.26). (d) and (e) Autocorrelation plots $C_{1-n_i}(\tau)$ defined in Eq. (7.28) for both zones. For Ancona the exponential fit has a decay rate of ≈ 11 days, while for Pescara the decay is of ≈ 33 days.

preliminary study of the data shows that there is a significative reduction of fishing activity from Friday to Sunday, consistent with markets being open Monday through Thursday only. The former was therefore dropped from the data set, keeping only trading days to ensure significant fishing activity.

7.2.2 *Defining fishing areas*

To assign each trawler with its base port (Ancona or Pescara), I used the following heuristic procedure, which was then cross-validated with MMSI data provided by the Ancona market authorities. The following notations are now introduced:

- $h^i(x,t)$ the time spent by trawler i fishing at grid-point x on day t,
- $w^i(x) := \sum_t h^i(x,t) / \sum_{y,s} h^i(y,s)$, for the average fraction of time spent by trawler i fishing at point x,
- $d_\mathrm{A}(x)$ the distance between point x and Ancona, and d the distance between the two cities,

- $d_{\mathrm{A}}^i := \sum_x w^i(x) d_{\mathrm{A}}(x)$, the average distance separating trawler i and Ancona when it is fishing,
- $D_{\mathrm{A}}^i := \sum_x w^i(x) [d_{\mathrm{A}}(x)]^2$, the average square distance between trawler i and Ancona.

and of course symmetrically for Pescara with index P. I then define the neighborhood of Ancona and Pescara as the pseudo-ellipsoid with focal points the two ports, i.e. the set $\{x \mid d_{\mathrm{A}}(x)^2 + d_{\mathrm{P}}(x)^2 \leq 2d^2\}$, of course excluding land, see Fig. 7.4(c). I restrict the analysis to trawlers evolving within this area, namely $\{i \mid D_{\mathrm{A}}^i + D_{\mathrm{P}}^i \leq 2d^2\}$. I've then assigned the trawlers to one of the two ports according to their average distance to each of them.

Defining two distinct areas as:

$$\begin{aligned} \mathscr{D}_{\mathrm{A}} &= \left\{x \mid d_{\mathrm{A}}(x) \leq d_{\mathrm{P}}(x) \quad \text{and} \quad d_{\mathrm{A}}(x)^2 + d_{\mathrm{P}}(x)^2 \leq 2d^2\right\} \\ \mathscr{D}_{\mathrm{P}} &= \left\{x \mid d_{\mathrm{P}}(x) < d_{\mathrm{A}}(x) \quad \text{and} \quad d_{\mathrm{A}}(x)^2 + d_{\mathrm{P}}(x)^2 \leq 2d^2\right\}, \end{aligned} \tag{7.26}$$

a given trawler is assigned to, say, Pescara if its fishing time-weighted average position lies in $\mathscr{D}_{\mathrm{P}}$. In other words $i \in$ Pescara (resp. Ancona) if $d_{\mathrm{P}}^i \leq d_{\mathrm{A}}^i$ (resp. $d_{\mathrm{A}}^i < d_{\mathrm{P}}^i$). This method of home port identification was validated by confronting it to the list of the Ancona-based trawlers, kindly provided by the Ancona fish market aurthorities. Up to a few minor errors, notably related to having identified as Ancona-based a few vessels based in the much smaller San Benedetto del Tronto, the cross-check was successful. Over the whole period, there are $N_{\mathrm{A}} = 108$ Ancona-based and $N_{\mathrm{P}} = 118$ Pescara-based trawlers.

7.2.3 *Stylized facts*

Having tagged each boat to either Ancona or Pescara, I now turn to studying the dynamics of fishing within the two areas $\mathscr{D}_{\mathrm{A}}$ and $\mathscr{D}_{\mathrm{P}}$. I define the fraction $n_{\mathrm{A}}(t)$ of time spent by Ancona-based vessels fishing in $\mathscr{D}_{\mathrm{A}}$ namely:

$$n_{\mathrm{A}}(t) = \frac{\sum_{x \in \mathscr{D}_{\mathrm{A}}, i \in \mathrm{Ancona}} h^i(x,t)}{\sum_{y, i \in \mathrm{Ancona}} h^i(y,t)}, \tag{7.27}$$

and vice-versa $n_{\mathrm{P}}(t)$ for Pescara. Figure 7.4(b) displays the evolution of $n_{\mathrm{A}}(t)$ and $n_{\mathrm{B}}(t)$ throughout the period of interest. While these fractions are most often very close to 1, indicating as one would intuitively expect that trawlers spend most of their time fishing near their home port, one can see that they regularly undergo persistent excursions, revealing that a sizeable fraction of the vessels in each area decide collectively to go elsewhere.

To evaluate the typical length of such excursions, Figs. 7.4(d) and (e) display the auto-correlation functions:

$$C_{1-n}(\tau) := \mathrm{Cor}\left(1 - n(t+\tau), 1 - n(t)\right), \tag{7.28}$$

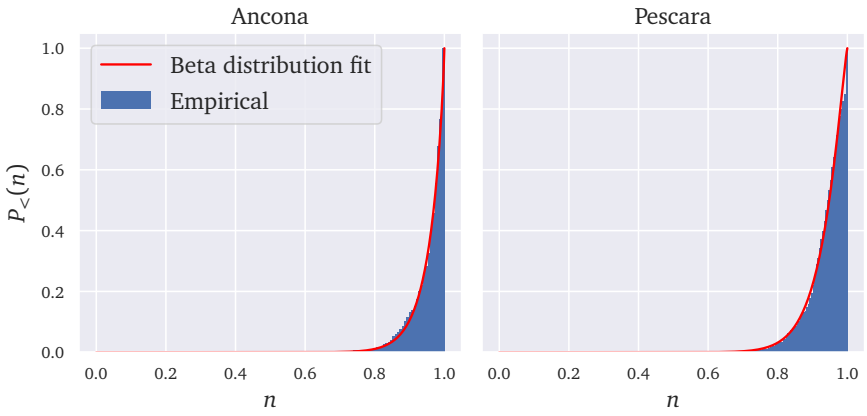


Figure 7.5: Cumulative distribution function (cdf) of the fractions n_A and n_P as defined in Eq. (7.27). The solid red curves correspond to a fit with a generalized Beta distribution, which has a cdf given by $P_>(n) = C\int_0^n \mathrm{d}x\; x^{\gamma_0-1}(1-x)^{\gamma_1-1}$ with C a normalization constant. The parameters for Ancona are $\gamma_0 = 18.48$ and $\gamma_1 = 0.82$, while those for Pescara read $\gamma_0 = 17.73$ and $\gamma_1 = 1.27$.

for both $n_A(t)$ and $n_P(t)$. These are well fitted by the sum of a delta-peak at 0, which can be attributed to measurement noise and other exogenous factors such as the weather,and an exponentially decaying function with typical time-scale ranging from ≈ 11 to ≈ 30 days. Interestingly enough, Fig. 7.5 reveals that the the empirical distributions of n_A and n_P are remarkably well fitted by a Beta distributions.This is exactly what one obtains in the ant model studied previously, in Sec. 7.1. In that model however, the two food sources are strictly equivalent and the resulting Beta distribution describing the fraction of ants at each source necessarily symmetric, at odds with the results obtained in the present setting. This motivates the asymmetric zones model introduced below. Another significant difference with Kirman and Föllmer's model is that the "food sources" here are not inexhaustible, as fishes do not have the ability to reproduce at in infinite rate.

These empirical results and observations motivate the introduction a model extending Kirman's original ant recruitment model to the present context. In essence, one can think of the two cities as two distinct ant colonies that can prey on any of the two zones. For each colony, the further fishing area is necessarily less attractive, allowing for the asymmetric character of the distribution. In addition, at odds with the ant model, this is not in a setting with unlimited resources and the new model should take into account the fact that over-fishing may deplete the sea.

7.2.4 *Description of the model*

The following is an extension to Kirman and Föllmer's original model that attempts to account for *exhaustible* and asymmetric sources, and that

attempts to reproduce the main stylized facts described in the previous section. Seeking to model fishermen exploiting a set of fishing areas, it imagines that boats follow the same basic dynamics as the ants: if they initially fish within a certain zone, they may decide to move elsewhere either because they see their peers fishing there, deciding to imitate them because they assume that their yield is good, or spontaneously decide to move elsewhere randomly for the sake of exploration.

This model has two major differences that depart from the original ant-recruitment model. First, it considers that a fishing area has finite resources: fish reproduce until reaching a certain finite capacity but they are also depleted by fishermen in the area (as in e. g. MacArthur's models [173, 172]). As a consequence, the random switching rate at which fishermen decide to depart from a given area depends on the fish population of that area. Note that this is very close in spirit to the modelling done in Ref. [6], but while also taking into account imitative behaviour in fishermen. The second difference with the ant model is that one imagines two "colonies" instead of just one, corresponding to vessels based at the two different fishing ports of Ancona and Pescara. Guided by the idea that fishermen prefer to go to areas close to their own home port, the model also introduces an asymmetry between the fishing areas for each food source.

The two ports, labelled A for Ancona and P for Pescara, have two distinct populations of fishermen, which may decide to exploit two fishing areas, S_1 and S_2, with the fishermen from A preferring to fish at S_1 and vice versa. One may of course reasonably argue that this view is far too coarse-grained, and that there may be, for example, many different fishing areas that are available close to each port. It is however possible to show under mild hypotheses that the two zones S_1 and S_2 in the model can be seen as the aggregation of a large number of smaller areas, with the same dynamics. For clarity, I shall define the model in discrete time, before moving into continuous time for analytical convenience.

Without loss of generality, I focus only on the dynamics of fishing vessels at one of the two ports, say Ancona, as I assume that fishermen only interact with boats coming from the same city[4]. Define now N_A and N_P as the number of boats based at Ancona and Pescara respectively, and let each of them decide to go to any of the two areas S_1 and S_2. I denote $m_i(t)$, with $i = 1, 2$, their respective fish populations at time t, and further assume that:

- Boats only fish in one area each day (consistent with discussions with port authorities) and come back to that area if they don't decide to switch to another one for the next day.
- A vessel's daily catch $c_i(t)$ is proportional to the amount of fish available in the area: $c_i(t) = \frac{\beta}{N_A} m_i(t)$ with $\beta / N_A \in [0; 1]$.[5]

4 Anecdotal evidence suggests indeed that the main interaction between people working in different boats happens at port in the fishing market or during informal conversation.

5 Without changing the main conclusions, one could also allow for noise by drawing $c_i(t)$ from a given distribution centred about $\beta m_i(t)/N$. This would allow introducing randomness

- Fish reproduce at a multiplicative rate ν_i, which I will take to be equal to ν for both areas.
- As a first approximation, fish do not travel from one area to the other.[6]
- The fish population within any area cannot exceed a carrying capacity K_i, which is the maximal population that can be present within an area in the absence of fishing. This carrying capacity is the same for all areas, as I have taken all of them to be equivalent. Without loss of generality, I take $K_1 = K_2 = 1$ in all that follows.

Note that these definitions, which also amount to thinking of the fish population as consisting of the same species in both areas, is justified by the fact that I only consider empirical data concerning trawlers. These boats fish only very specific, shallow water dwelling species.

I further define $N_{A,i}(t)$ the number of vessels from port A fishing at zone i at time t (and $N_{P,i}(t)$ respectively). The number of fishing vessels in each port is fixed, implying for all t: $N_{A,1}(t) + N_{A,2}(t) = N_A$. The different assumptions translate into following evolution for the fish population:

$$m_i(t+1) - m_i(t) = m_i(t)\left[\nu g(m_i(t)) - \beta\left(N_{A,i}(t) + N_{P,i}(t)\right)\right], \quad (7.29)$$

where the function g must satisfy $g(0) = 1$ and $g(1) = 0$.

The simplest assumption one can make is that of logistic growth, leading to $g(m_i(t)) = 1 - m_i(t)$, as in Eq. (5.12). It follows that $m_i(t) \in [0;1]$, $\forall t$, where $m = 1$ corresponds to a fishing area at full capacity and $m = 0$ corresponds to a depleted area. Under these assumptions, the evolution of the fish population is of the Lotka-Volterra type, as advocated in [6][7].

Furthermore, I've assumed that a fishing vessel based at A fishing at i can randomly decide to go elsewhere with probability $\varepsilon_{A,i} f(m_i(t))$, where the function f satisfies $f(1) = 1$ and $f(0) = 1 + \kappa$. Here, $\varepsilon_{A,i}$ controls the base intensity of the noise, that can take a maximal value $\varepsilon_{A,i}(1+\kappa)$ when the zone is depleted. Fishermen have then a higher incentive to go elsewhere as their fishing yield decreases, and the preference of fishermen from A for zone 1 is highlighted by setting $\varepsilon := \varepsilon_{A,1} = \varepsilon_{A,2}/C_d$ with $C_d > 1$

into the fishing efficiency of each trawler, an interesting extension that is left to further work.

6 This constraint can be easily relaxed by e.g. adding a migration term where fishes from 2 move to 1 at a certain rate and vice-versa. In practice, this would only tend to prevent the difference between the two fish populations from fluctuating too wildly.

7 As an interesting anecdote, I learned in [27] that "Vito Volterra was born in the Jewish ghetto of Ancona in 1860, shortly before the unification of Italy, when the city still belonged to the Papal States", and that "in 1925, at age 65, Volterra became interested in a study by the zoologist Umberto D'Ancona, who would later become his son-in-law, on the proportion of cartilaginous fish (such as sharks and rays) landed in the fishery during the years 1905–1923 in three harbours of the Adriatic Sea: Trieste, Fiume and Venice. D'Ancona had noticed that the proportion of these fish had increased during the First World War, when the fishing effort had been reduced". This led him to take interest in models that Alfred Lotka had first used to model very general population dynamics, and that I now apply, without knowing any of this at first, to the fish population dynamics at the ports of Ancona and Pescara.

a parameter controlling the degree of asymmetry between zones S_1 and S_2 for a fisher from A. This allows to have a larger spontaneous switching rate $S_2 \to S_1$ for fishermen from A.

Besides this random switching rate, the model adds a crucial element, which is that agents imitate each other. Each day, a fisher randomly picks one of his peers at random and decides to imitate him/her with probability μ/N, so that μ is the fraction of boats deciding to take an imitation strategy at each step. In this case, the probability that a boat from A initially at zone S_i decides to move to zone S_j is given by:

$$P_A(S_i \to S_j) = \varepsilon_{A,i} f\left(m_i(t)\right) + \frac{\mu}{N_A} \frac{N_{A,j}(t)}{N_A - 1}, \tag{7.30}$$

similar to Eq. (7.2).

As with the ant model, I define $n_{A,i} = N_{A,i}/N_A$. Taking the limit $N_A, N_P \to \infty$ with $N_P/N_A = C_N$ fixed, it is possible to write the following Fokker-Planck equation for the probability density $\rho(\mathbf{n}_A, \mathbf{n}_P, \mathbf{m})$, where $\mathbf{n}_A = (n_{A,1}, n_{A,2})$ and $\mathbf{m} = (m_1, m_2)$, but this is in practice insolvable.

The key lies in doing a mean-field approximation. Take the continuous version of Eq. (7.29),

$$\frac{\mathrm{d}m_1}{\mathrm{d}t} = m_1(t)\left(\nu(1-m_1(t)) - \beta\left(n_{A,1} + C_N\left(1 - n_{P,2}\right)\right)\right), \tag{7.31}$$

and set $\mathbb{E}\left[\frac{\mathrm{d}m_1}{\mathrm{d}t}\right] = 0$ to get the condition:

$$\mathbb{E}\left[m_1\right] = \left[1 - \frac{\beta}{\nu}\left(\mathbb{E}\left[n_{A,1}\right] + C_N\left(1 - \mathbb{E}\left[n_{P,2}\right]\right)\right)\right]_+, \tag{7.32}$$

where $[x]_+ = x\mathbf{1}_{x>0}$ denotes the positive part of x. In particular, one can see that there exists an extinction line for the fish population for:

$$\nu = \beta\left[\mathbb{E}\left[n_{A,1}\right] + C_N\left(1 - \mathbb{E}\left[n_{P,2}\right]\right)\right], \tag{7.33}$$

corresponding to the case where the reproductive rate of fish corresponds exactly to the rate at which they are fished.

Insert then Eq. (7.32) into the vessels' dynamics of Eq. (7.30) by replacing the argument of $f(m_i)$ by the average, as $f\left(\mathbb{E}\left[m_i\right]\right) := f_i$. Choosing, for the sake of definiteness, a linear function for f, i.e. $f(x) = 1 + \kappa(1-x)$, the average $\mathbb{E}\left[n_{A,1}\right]$ can now be easily computed as:

$$\mathbb{E}\left[n_{A,1}\right] = \frac{C_d f_2}{C_d f_2 + f_1} = \frac{C_d\left(1 + \kappa(1 - \mathbb{E}\left[m_2\right])\right)}{2 + \kappa\left[1 - \mathbb{E}\left[m_1\right] + C_d\left(1 - \mathbb{E}\left[m_2\right]\right)\right]}. \tag{7.34}$$

7.2.5 *Stationary solutions*

Within this mean-field approximation, I set $m_1 = \mathbb{E}\left[m_1\right]$ (resp. $m_2 = \mathbb{E}\left[m_2\right]$) and use the same methods as those of the previous Sec. 7.1 to

obtain the following Fokker-Planck equation for the density $\rho\left(n_{A,1}, n_{P,2}, t\right)$ of the quantities $n_{A,1}$ and $n_{P,2}$:

$$\partial_t \rho = \partial_{n_{A,1}} J_1 + \partial_{n_{P,2}} J_2, \tag{7.35}$$

where:

$$J_1 = -\varepsilon\left[C_d f_2 - n_{A,1}\left[C_d f_2 + f_1\right]\right]\rho + \mu\partial_{n_{A,1}}\left[n_{A,1}(1-n_{A,1})\right]\rho, \tag{7.36}$$

and where the transposition to find the definition of J_2 is transparent.

As for the ant model, the stationary state is found by setting $J_1 = 0$ and $J_2 = 0$ and solving for ρ. The mean-field approximations decouples the evolution of the two variables and allows one to write:

$$\rho(n_{A,1}, n_{P,2}) = \rho_1\left(n_{A,1}\right)\rho_2\left(n_{P,2}\right), \tag{7.37}$$

with:

$$\rho_1(n_{A,1}) = C_1 n_{A,1}^{\gamma_{A,0}-1}\left(1-n_{A,1}\right)^{\gamma_{A,1}-1}, \quad \rho_2\left(n_{P,2}\right) = C_2 n_{P,2}^{\gamma_{P,0}-1}\left(1-n_{P,2}\right)^{\gamma_{P,1}-1}, \tag{7.38}$$

where C_1 and C_2 are normalisation constants and the γ parameters for the ρ_1 distribution (the parameters for ρ_2 can be easily deduced) read:

$$\gamma_{A,0} = \frac{\varepsilon}{\mu} C_d f_2, \quad \gamma_{A,1} = \frac{\varepsilon}{\mu} f_1. \tag{7.39}$$

Note also that full dynamical solutions $\rho(n_{A,1}, t)$, $\rho(n_{P,2}, t)$ can be obtained in terms of hypergeometric functions, in the same spirit of Ref. [189] and as described in Sec. 7.1.

The model thus replicates successfully the observed distributions shown in Figure 7.5, and captures the qualitative behaviour from Figure 7.4. An example of simulation from the model is provided in Figure 7.6. For this Figure, I have set $\mu = 1$ as it only amounts to a certain choice of the time-scale, while picking a small value $\kappa = 0.1$ to keep f_2, $f_1 \approx 1$ and I have then picked ε and C_d as to obtain the values for γ_0 and γ_1 from Figure 7.5.

Note also that, as with the ant model, it is possible to compute the auto-correlation of $1-n_{A,1}$, defined as in Eq. (7.28), to find

$$C_{1-n_{A,1}}(\tau) = \exp\left(-\mu\left(\gamma_{A,0} + \gamma_{A,1}\right)\tau\right), \tag{7.40}$$

which is exactly what one sees from the data in Fig. 7.4 (d) and (e), provided one interprets the delta-peak at $\tau = 0$ as the result of exogenous noise, e.g. weather conditions.

Indeed if one considers that the measured signal is in fact a noisy signal,

$$\tilde{n}(t) = (1-\sigma)n(t) + \sigma\xi(t), \tag{7.41}$$

where $n(t)$ is the "true" process and $\xi(t)$ is a gaussian white noise of unit one, then one can show directly that the measured correlation function reads

$$C_{1-\tilde{n}}(\tau) = \delta(\tau) + \frac{(1-\sigma)^2}{\sigma^2} C_{1-n}(\tau). \tag{7.42}$$

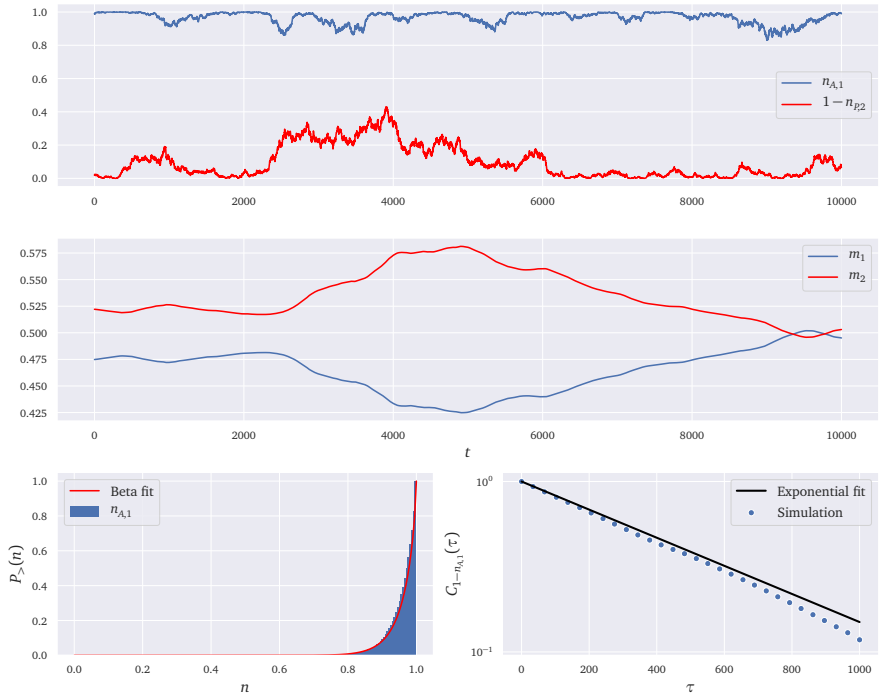


Figure 7.6: A simulation of the model. I have chosen the different parameters as to obtain the same stationary Beta distribution as observed in Fig. 7.5. Note also the similarity of the Figure on the left with plot (a) in Figure 7.4. I have used the parameters $\varepsilon = 0.69$, $\nu = 10$, $\beta = 5$, $C_N = 1$, $C_d = 26.5$ for Ancona and 15.7 for Pescara and $\kappa = 0.1$. The upper panel shows the two trajectories $n_{A,1}(t)$ and $n_{P,2}(t)$ and the middle panel shows the fish populations $m_1(t)$ and $m_2(t)$, while the bottom left panel shows the cumulative density function for $n_{A,1}$ along with a Beta distribution fit, and the bottom right panel shows the empirical correlation function as defined by Eq. (7.40) along with an exponential fit. Note that the fish populations oscillate around the theoretical mean-field value $m = 0.5$, and that large oscillations coincide with large collective movements of the fishermen in both areas.

More complicated correlation functions can also be computed, along the lines of Sec 7.1, although they are more prone to statistical noise. For example, one can show that the polynomial defined by

$$\sigma_A(n_{A,1}) = n_{A,1}^2 - \frac{2(\gamma_{A,0}+1)}{\gamma_{A,0}+\gamma_{A,1}+2} n_{A,1} \tag{7.43}$$

has an autocorrelation function that is exponential, meaning that $C_{\sigma_A}(\tau) = \mathrm{Cor}\big(\sigma_A\big(n_{A,1}(t+\tau)\big), \sigma_A\big(n_{A,1}(t+\tau)\big)\big)$ verifies

$$C_{\sigma_A}(\tau) = \exp\big(-2\mu(1+\gamma_{A,0}+\gamma_{A,0})\,\tau\big), \tag{7.44}$$

and the same definition can of course be transposed to the variables indexed by P.

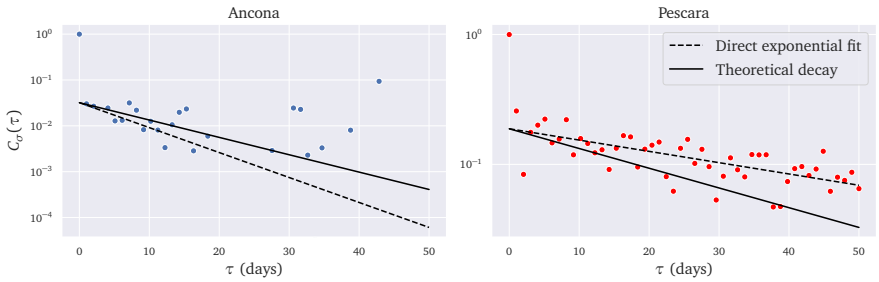


Figure 7.7: Empirical correlation function $C_\sigma(\tau)$ as defined in Eqs. (7.43) and (7.44). The solid black line is the theoretical prediction given the estimations of γ_0 and γ_1 from the empirical probability distribution in Figure 7.5 and from the subsequent estimation of μ using the exponential decay factor from Figure 7.4. The reliable computation of σ depends of course on the proper estimation of these parameters, and one would expect them to be noisy. Nonetheless, the agreement with theory, especially in the case of Pescara, is good.

The results of this prediction along with those given by the data are shown in Fig. 7.7. This correlator is necessarily more affected by noise, because it is of order two in the n variables and because it depends on a reliable estimation of the γ and μ variables. Despite these limitations, the theoretical prediction is satisfactory when compared with the data, especially in the case of Pescara.

7.2.6 *The symmetric limit*

In general, the fixed point Eq. (7.32) linking the averages $\mathbb{E}[m_i]$ with the averages $\mathbb{E}[n_1]$ cannot be solved directly. Nonetheless, if one takes $C_N = 1$ to have completely symmetric fishing areas, then the equations simplify considerably as this immediately implies $f_1 = f_2$, with then Eq. (7.34) becoming

$$\mathbb{E}\left[n_{A,1}\right] = \mathbb{E}\left[n_{P,2}\right] = \frac{C_d}{C_d + 1}. \tag{7.45}$$

One can then write Eq. (7.32) more explicitly, to obtain the following extinction line,

$$\mathbb{E}[m_1] = \mathbb{E}[m_2] = \begin{cases} 1 - \frac{\beta}{\nu} & \text{if} \quad \beta < \nu \\ 0 & \text{if} \quad \beta \geq \nu \end{cases}, \tag{7.46}$$

which has the intuitive interpretation that the population within a given area goes extinct if the fishing rate is larger than the reproduction rate of the fish there. I show in Figure 7.8 that the agreement of simulations with the mean-field analysis is excellent. One should note however that this convergence may be slow as $\nu \to 0$, as this parameter controls the global time-scale of the fishes.

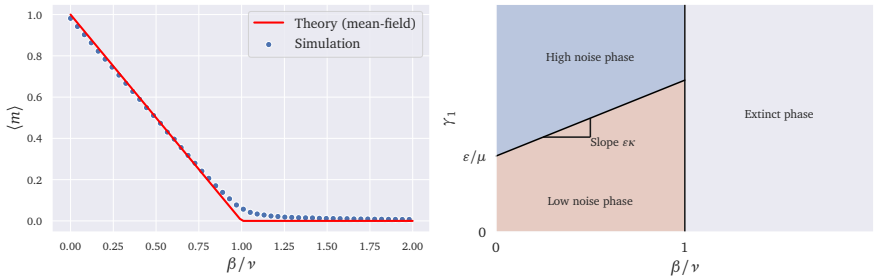


Figure 7.8: Left: simulation results with the same parameters as previously. The simulation was run for $T = 10^5$ steps and with $\nu = 10$. Note that the convergence of the simulation to the mean-field results from Eq. (7.46) is better as T or ν grow larger. Right: phase diagram of the model.

Note also that in this case, one can directly compute $f_1 = f_2 = 1 + \frac{\kappa\beta}{\nu}$. In this case, the parameters in Eq. (7.39) simplify to yield

$$\gamma_0 = \frac{\tilde{\varepsilon}}{\mu} C_d, \quad \gamma_1 = \frac{\tilde{\varepsilon}}{\mu}. \tag{7.47}$$

where I've dropped the A index as the parameters for both areas A and P are identical, and where I've set $\tilde{\varepsilon} = \varepsilon\left(1 + \frac{\kappa\beta}{\nu}\right)$.

In this limit it is then clear that the mean-field model amounts to a modification of the original ant model, where the noise ε is augmented because of the sensitivity of the fishermen to the local fish population by the factor given above, and where there is an asymmetry between the two areas/food-sources through the parameter C_d.

One would then typically expect that $\gamma_0 > 1$ always because of the strong preference for the fishing area closest to one's port. However, if ε or κ are strong, one can have a crossover at $\gamma_1 = 1$. The simulations on Figure 7.6 correspond both to $\gamma < 1$, the empirical data shown in Figures 7.4 and 7.5 has $\gamma_1 < 1$ for Ancona, and $\gamma_1 \gtrapprox 1$ for Pescara. When $\gamma_1 > 1$ the behaviour is qualitatively different: instead of having the majority of the boats nearly always fish at the closest area, with occasional "jumps" to go to the neighbouring zone, there is always a degree of "mixing", as at any given time there is always a fraction $\approx 1 - \mathbb{E}\left[n_{A,0}\right]$ of fishermen from Ancona fishing near Pescara. I show a simulation of this case, with γ_1 well above 1, in Figure 7.9.

7.3 CONCLUSION

In the last Section, I have shown the possible application of the ant recruitment model to more broad settings. The extended model takes into account the possibility of having asymmetric, non-exhaustible sources and shows very good qualitative agreement with the main features present in empirical data by producing Beta distributions and exponentially decaying correlations. This could have been enhanced had access to detailed from

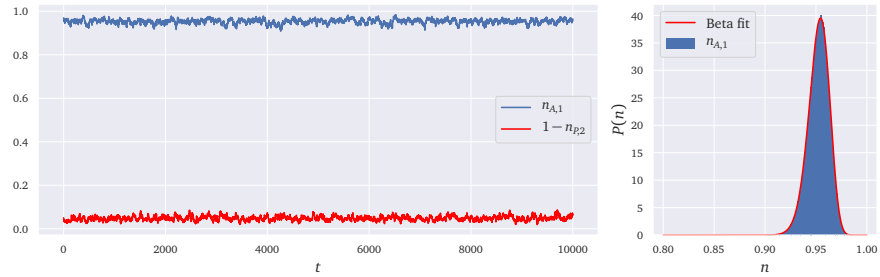


Figure 7.9: Plot of a simulation in the case $\gamma_1 > 1$. The parameters are the same as that of Figure 7.6, but with $\kappa = 10$ instead. The fit gives $\gamma_0 = 220$ and $\gamma_1 = 20$, while the predicted theoretical values from Eq. 7.47 are $\gamma_0 = 220$ and $\gamma_1 = 11$. Note that, contrary to Figure 7.5, I've depicted the density directly instead of the cumulative density function.

the fish market been possible, but a visit to the Ancona and Pescara markets was cut short because of the Covid crisis. This would of course have allowed to validate some of the hypotheses of the link between the boats' dynamics and the fish population, which could have been estimated from the amount of fish sold at the markets.

I believe however that the generic behaviour of the model, due mostly to the noise term being proportional to $\sqrt{x(1-x)}$, is quite generic, as shown by its usefulness in describing different processes in chemistry, genetics or in the voter model. The resolution technique given in Section 7.1 can therefore have applications much beyond the setting exposed here, but I think that one of the interesting results that it allows is that it allows to compute precisely the *ergodic time* $1/\varepsilon$.

This means, for example, that if one observed the behaviour of the ants for a time shorter than $1/\varepsilon$ then one could easily think that one food source is preferred over the other, because one would then be studying the system for a time much shorter than the time it takes to reach *statistical* equilibrium. This is an hypothesis that is often made in economics, and that allows for example to replace cross-sectional (i. e. across all agents/entities) averages with a time-average, viz. one that is made for a single agent or entity after observing it for a very long time. As shown in the study of the Random Energy Model in Sec. 3.3.3.1 this is not always the case, and an argument has been made recently (see e. g. Ref. [212]) to drop the ergodicity hypothesis and develop models that go beyond it. This is one of the issues dealt with in the next Chapter, where I will show that memory effects can lead to very non-trivial dynamics and to a form of *weak ergodicity breaking* with interesting links to the statistical physics of disordered systems.

8

BY FORCE OF HABIT

In modelling the behaviour of agents, the common view in standard economics is that their actions are guided by the maximization of an utility function. For convenience, the utility for each individual is often thought of as independent of the actions of others, as well as static in time, and a myriad of results in economic theory actually rest upon this assumption. Complexity economics has recently begun to tackle the issue of interactions between agents with analytical and numerical tools, and I address here the possibility of reinforcement mechanisms that make an agent's utility depend on his past.

In rational choice theory, individuals set their preferences according to an utility maximization principle. Each choice an individual can make is assigned a certain "utility", i. e. a quantity measuring the satisfaction it provides to the agent and frequently related to the dispassionate forecast of a related pay-off. This framework is often accompanied by the assumption that the agent considers *all* available choices present to her/him, weighs their utilities against one another, and then makes her/his choice taking into account possible constraints, such as a finite budget.

A number of criticisms to this view of human behaviour have emerged, with e.g. Simon [241] as a key figure highlighting that individuals may be "satisfiers" rather than pure optimisers, in the sense that there is both a computational cost and a cognitive bias related to considering the universe of available choices. Sometimes finding the optimum of the utility function can itself be such a computationally hard problem that even the most powerful computers would not be able to find it in a reasonable amount of time. This led to the idea of bounded rationality as a way to model real agents [243, 233, 22, 129].

More recently, Kahneman [148, 149] pointed at what he considers to be significant divergences between economics and psychology in their assumptions of human behaviour, with a special emphasis on the empirical evidence of the cognitive biases, and therefore the *irrationality*, that guides individual behaviour. A pervasive effect, for example, is that the utility of a certain choice strongly depends on the choice made by others. These so called "externalities" can lead to interesting collective effects, where choices made by agents synchronise and condense on a small subset of choices, or lead to confidence crises – see for example [69, 104, 63, 194, 53].

An interesting idea developed in [147] is the fact that the utility associated to a certain decision may depend also on our memory if it has already been made in the past. The model proposed below encapsulates this idea, and I will show that this too can lead to choices that do not necessarily

conform to their "objective" utilities, but are rather dominated by past choices alone. This is related to what economists call "habit formation" [78, 1, 92, 79, 116, 219]. Memory effects chisel the utility landscape in a way that may render objectively sub-optimal choices subjectively optimal. In the case of sufficiently long range memory, agents may, in a self-fulfilling kind of way, become "trapped" forever in a certain choice and stop exploring alternative choices. Moreover, this self-fulfilling mechanism completely breaks ergodicity.

The model below, done with A. Fosset, D. Luzzati, M. Benzaquen and JP Bouchaud in Ref. [192] and inspired from [63], assumes that the utility landscape is affected by past choices, with a memory kernel that decays with time. Agents can change their decision using a logit (or Metropolis) rule, parametrised by an "intensity of choice" β that plays the role of the inverse temperature in statistical physics. This type of model belongs to a wide class of so-called "reinforcement" models, which contains Polya Urns, Reinforced Random Walks, Elephant Walks, etc. – for a review see [209] and references therein.[1] Such models have also gained traction in the economics literature, where positive reinforcement of certain choices made by agents are shown to impact the emergence of certain macro outcomes and structures [23, 21].

8.1 A SIMPLE MODEL FOR DECISION WITH MEMORY

Consider a set of N discrete choices, labelled $(x_i)_{1\leq i\leq N}$, to which one assigns an utility – a measure of the value an individual assigns a given choice. The perceived utility of site x_i and time t is postulated to be:

$$U(x_i,t) = U_0(x_i)\left(1+\sum_{t'=0}^{t}\phi(t-t')\mathbf{1}_{x(t')=x_i}\right), \tag{8.1}$$

where the first term on the right-hand side is the intrinsic, or objective utility of the choice, while the second accounts for memory effects, affecting the utility of that choice for the only reason that the individual has picked it in the past.[2] The decaying memory kernel ϕ encodes that more recent choices have a stronger effect, and $x(t)$ denotes the choice of the individual at time t. Hence, past history "chisels" the utility landscape, in a way similar to ants leaving a pheromone trace that guide other ants along the same path, or rivers creating their own bed through erosion. Note that in most reinforcement random walk models reviewed in [209], infinite memory span is assumed, i. e. $\phi(t) =$ constant, while I will be mostly concerned here with decaying memory kernels.

The sign of the kernel ϕ separates two different cases: $\phi < 0$ indicates a situation where an individual grows weary of his past choices, while

1 For recent developments, see also [68, 146].

2 One may also think, in the physicist's language, of an energy landscape (akin to minus the utility) where the energy of a given site or configuration increases or decreases if the system has already visited that site.

$\phi > 0$ corresponds to the case where an individual becomes increasingly endeared with them. In agreement with intuition, the former case leads to an exploration of all the choices unless the optimal choice has an utility too far apart from the rest to be sufficiently affected by the kernel.

In all that follows I focus on the more interesting case $\phi \geq 0$. The reason behind studying such utility reinforcement lies in the behavioural idea that people tend to prefer what they already know, thus paving the way for "habit formation" as in [156], see also [1, 92, 79, 116, 219] and [63].

I now invite the reader to consider the following dynamics. An individual, standing by choice x_i at time t, draws an alternative x_j from a certain ensemble of "nearby choices" ∂_i, e.g. the set of neighbours of i in a graph $\mathcal{G}$, with probability:

$$\mathcal{T}_{x_i \to x_j} = \frac{\mathbf{1}_{x_j \in \partial_i}}{\mathcal{N}_i}, \quad \mathcal{N}_i := \sum_j \mathbf{1}_{x_j \in \partial_i}, \tag{8.2}$$

where $\mathcal{N}_i$ is the number of neighbours of i. Restricting to nearby choices is a parsimonious way to model adaptation costs, that penalize large decision changes. However, the framework I am presenting is quite versatile since the topology of the graph $\mathcal{G}$ is arbitrary, and I will consider different cases below.

The target choice x_j is then adopted with the logit probability, standard in Choice Theory [18]:[3]

$$p(x_i \to x_j) = \frac{1}{1 + e^{\beta[U(x_i,t) - U(x_j,t)]}}, \tag{8.3}$$

where β is called the "intensity of choice" and accounts for the degree of rationality (it is the analogue of the inverse temperature in statistical mechanics). Indeed, as long as $0 < \beta < \infty$, the agent is more likely to switch whenever $U(x_j, t) > U(x_i, t)$ (optimizing behaviour), but the probability to pick a choice with a lower utility is non-zero, which encodes for bounded-rationality (or uncertainty about the true utility) in the economics literature. In the $\beta \to 0$ limit (equivalent to the infinite temperature limit in physics) the agent explores the whole space of possible choices without taking their utility into account. In the opposite limit $\beta \to \infty$ (or zero temperature) the agent has a greedy behaviour and only switches to choices with a higher utility, but this also implies that he/she may stay in a local maximum instead of taking the chance to explore all available possibilities. An illustration of these dynamics is given in Fig. 8.1.

When $\phi = 0$, the dynamics is that of the Metropolis-Hastings algorithm used to sample the Boltzmann-Gibbs distribution [182, 140]. The stationary state of the dynamics is such that the probability to pick choice x_i is $p_i \propto \mathcal{N}_i e^{\beta U_0(x_i)}$, as one can show easily by solving the detailed balance condition

$$p(x_i \to x_j) p_i = p(x_j \to x_i) p_j. \tag{8.4}$$

3 For non-trivial trapping to emerge, the model considers graphs without singletons, that is to say that all sites have a non-empty set of neighbours that are different to itself.

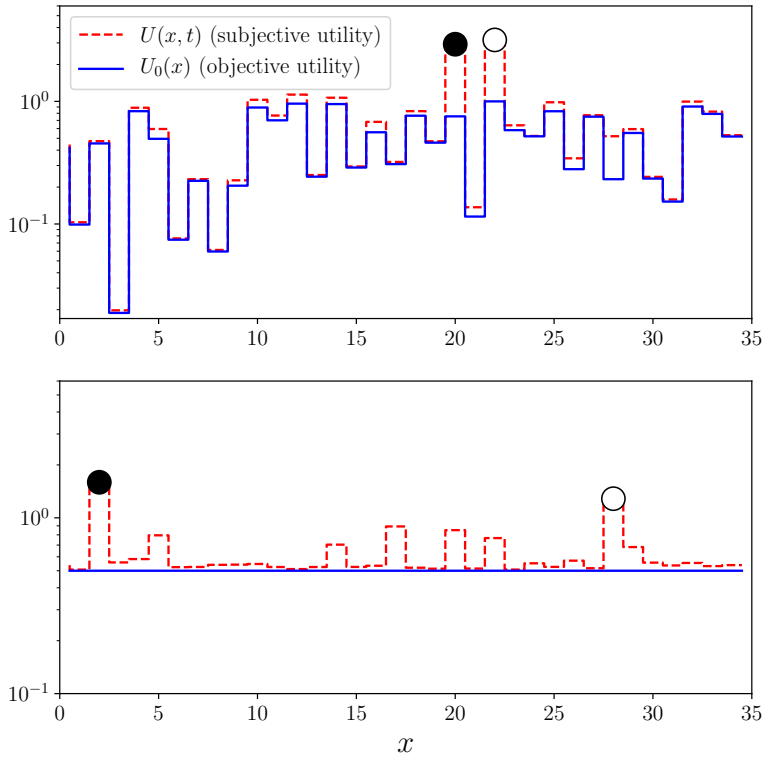


Figure 8.1: Schematic representation of the problem at a given time t. The plot on the top depicts the case of random "objective" utilities $U_0(x)$, while the one on the bottom shows the situation where they are uniform $U_0(x) = U_0$. In both plots, the solid black ball represents the choice made at time t, while the empty ball represents the choice made at time $t-1$. Both correspond to a simulation run with a power-law kernel $\phi(t) \propto (1+t)^{-\gamma}$ with $\gamma = 1.5$ and $\beta = 0.2$, on a fully connected graph.

This can by itself lead to interesting phenomena depending on the statistics of U_0. For example the study of the Random Energy Model [100, 65] as done in Sec. 3.3.3.1 shows that for a finite value of N and for Gaussian utilities of variance σ^2„ the probability measure condenses on a small number of choices, much smaller than N, for $\beta > \beta_c = \sqrt{2\ln N}/\sigma$.

8.2 NON-ERGODICITY & CONDENSATION OF CHOICES

Adding the kernel introduces the possibility that the agent gets stuck in a non-optimal choice exclusively through memory effects: staying a long time in a given choice self-reinforces its utility, thereby increasing the likelihood to stay there and leading to non-ergodic dynamics. To study the possible condensation or trapping induced by memory alone, I restrict the study to

the case where $U_0(x_i) = 1,\ \forall i$. The interplay between memory-induced trapping and utility heterogeneity is quite interesting in itself, but is left to future investigations.

I now consider an agent starting from a given choice x_0 at time $t = 0$ and follow his/her evolution for times $t = 1, \dots, T$ with T sufficiently large. I then compute the empirical state histogram $p_i = \sum_t \mathbf{1}_{x(t)=x_i}/T$ and define the order parameter h, the same as the herfindahl index defined in Sec. 2.3.1, as:

$$h := \sum_{i=1}^{N} p_i^2. \tag{8.5}$$

This parameter indicates how the agent has explored the space of possible choices: if all choices were visited with equal probability then one has immediately $p_i = 1/N$ and thus $h = \sum_i 1/N^2 = 1/N$. On the other hand if the agent was stuck in a single choice j, then $p_i = \delta_{i,j}$ and so $h = 1$. Therefore $1/h$ gives an order of magnitude of the number of different choices picked by the agent during time T. In practice, I average h over a large number of trajectories and starting sites x_0, to obtain an average parameter $\mathbb{E}[h]$. For a set of simulations on a graph $\mathcal{G}$ with N choices and lasting a time T, I therefore define the critical value β_c, defining the crossover between $h = \mathcal{O}(1/N)$ and $h = \mathcal{O}(1)$ as the value for β that maximizes the variance of h over different trajectories.

An important question is whether β_c corresponds to a true transition or to a mere crossover. This depends on the L^1 norm of the memory kernel, $\|\phi\| = \sum_{t=0}^{\infty} \phi(t)$. Suppose that this norm is finite. Then if the agent has been stuck in a given site i for a time $t \gg 1$, I can approximate its utility by $U_0(1 + \|\phi\|)$. The difference in utility with the neighbouring choices thus remains finite. For any finite value of β, the probability to leave that site is non-zero, and therefore the individual will eventually pick a different choice. The time for this to happen is however of the order of $\exp(\beta U_0 \|\phi\|)$. If this time is much longer than T, the experiment will yield $h \sim 1$, even though running the trajectory for a longer time would result in $h = \mathcal{O}(1/N)$. Hence in this case β_c is a crossover that depends on T as $\ln(T)/(U_0\|\phi\|)$.

A more interesting situation (at least from a theoretical point of view) is when $\|\phi\| = \infty$. As I will show below, there exists cases where β_c corresponds to a true phase transition and is independent of T (when T is large).

8.3 MEAN FIELD APPROXIMATION

In order to draw further analytical features, I start by looking at a mean-field approximation. This means totake the graph $\mathcal{G}$ to be fully connected with $\mathcal{T}_{x_i \to x_j} = 1/(N-1)$ and in the limit $N \to \infty$.

I will now formalize the argument previously sketched. If the individual started first at a given choice corresponding to node i, then the probability

$P_>(\tau)$ that he/she remains there up to a time τ is given by the product over $t \in [\![0, \tau-1]\!]$ of the probabilities not to leave the site between times t and $t+1$, $p_{\text{stay}}(t)$. Now, $p_{\text{stay}}(t) = 1 - p_{\text{leave}}(t)$ with:

$$p_{\text{leave}}(t) = \sum_{j \in \partial_i} \mathscr{T}_{x_i \to x_j} \frac{1}{1 + e^{\beta[U(x_i,t) - U(x_j,t)]}}. \tag{8.6}$$

For the fully connected graph, this expression simplifies to[4] $p_{\text{stay}}(t) = \left[1 + e^{-\beta\Phi(t)}\right]^{-1}$, with $\Phi(t) = U_0 \sum_0^t \phi(s)$. It follows that:

$$P_>(\tau) = \prod_{t=0}^{\tau} \left[1 + e^{-\beta\Phi(t)}\right]^{-1} \approx e^{-I(\tau)}, \tag{8.7}$$

with $I(\tau) := \int_0^\tau \mathrm{d}t \, \ln\left[1 + e^{-\beta\Phi(t)}\right]$, where discrete sums have been replaced by integrals. Equation (8.7) determines the distribution of the "trapping" time τ that the agent spends stuck on a certain choice. Its nature will entirely depend on the behaviour of the integral $I(\tau)$ when $\tau \to \infty$.

8.3.1 *Short Term Memory*

Consider first the case where $\lim_{t\to\infty} \Phi(t) = \|\phi\| < +\infty$. Then $I(\tau) \approx \lambda\tau$ for $\tau \to \infty$, with

$$\lambda := \ln\left[1 + e^{-\beta\|\phi\|}\right]. \tag{8.8}$$

This means that the trapping time distribution decays exponentially fast for large τ, with an average trapping time $\langle\tau\rangle$ approximately given by $1/\lambda$. For sufficiently small λ, one recovers the qualitative criterion of the previous section by setting $T\lambda \sim 1$. But the dynamics remains ergodic when $T \to \infty$.

8.3.2 *Long Term Memory*

Suppose now that $\phi(t)$ decays sufficiently slowly for large t for $\|\phi\|$ to diverge. For definiteness, I will focus on power-law kernels:

$$\phi(t) = \frac{C}{(1+t)^\gamma}. \tag{8.9}$$

When $\gamma > 1$, $\|\phi\|$ is finite and we are back to the previous case. Hence I restrict to $\gamma \leq 1$. When $\gamma < 1$, one finds that $\Phi(t) \propto t^{1-\gamma}$ for large t. Hence $I(\tau)$ converges to a finite limit I_∞ for large τ. This means that there is a finite probability $P_\infty = e^{-I_\infty}$ that the choice is made *forever*. When $\gamma = 1$, $\Phi(t) \approx CU_0 \ln t$ for large t. This leads to three further sub-cases:

4 In the general case in which the agent started in a "site" different from i and then got stuck in i, one wants to replace the right-hand side by an average of the logit rule over the utility gaps $U(x_i,t) - U(x_j,t)$. However, as $N \to \infty$ it is very unlikely that a site that was previously picked is chosen again. The reader should agree that one can thus safely replace the average gap by the gap with the base level $U_0 = 1$.

1. When $\beta C U_0 > 1$, $I(\tau)$ again converges to a finite limit when $\tau \to \infty$, i. e. decisions self-trap forever.

2. At the transition point, defined as $\beta_c^\star = (CU_0)^{-1}$, one finds that $P_>(\tau)$ decays as τ^{-1}, i. e. the trapping time distribution is a Zipf law, τ^{-2}. This is the marginal case that appears in several models of aging in the literature [45, 14]. For a finite observation time T, the average trapping time grows like $\ln T$.

3. When $\beta C U_0 < 1$, $I(\tau)$ behaves for large τ as $\exp\big(-\tau^b/b\big)$, where $b = 1-\beta C U_0 > 0$. The average trapping time $\langle\tau\rangle$ is thus finite. A careful analysis shows that $\langle\tau\rangle$ diverges as b^{-1} when $b \to 0$, but higher moments $\langle\tau^k\rangle$ with $k > 1$ diverge much faster, as $\exp((k-1)/b)$, i. e. according to the so-called Vogel-Tamman-Fulcher law, see e.g. [97].

To summarize, when the kernel ϕ decays fast enough, there is a crossover regime in β between free exploration of the space of choice and trapping. The crossover value of β depends on the observation time T and is given, using Eq. (8.8) for T large, by

$$\beta_c = \frac{\ln T}{\|\phi\|}. \tag{8.10}$$

When memory is long ranged, and described by a power-law kernel with decay exponent γ, there exists a genuine transition when $\gamma = 1$ between a free exploration regime and a (non-ergodic) trapped regime at a T independent value of β that I henceforth call $\beta_c^\star$.

When $\beta > \beta_c^\star$ or $\gamma < 1$, there is a non-zero probability to get trapped in the same decision forever. The characteristic time for changing decision is of the same order of magnitude as T itself, a phenomenon called "aging", see e.g. [59, 14], on which I will comment further below, see Fig. 8.4. Note finally that for $\gamma = 1$, the mean-field analysis predicts that while the average trapping time diverges as $(\beta_c^\star - \beta)^{-1}$, all higher moments diverge much faster, as $\sim \exp\big(A/(\beta_c^\star - \beta)\big)$.

8.4 NUMERICAL RESULTS

I conducted simulations using a long-range memory kernel given by Eq. (8.9) with $\gamma \in [1;\infty[$. For numerical convenience, I represented $\phi(t)$ as a superposition of exponentials as done in [51], and basing myself on code originally written by A. Fosset and D. Luzzati that I improved by translating it onto C++. I then considered a variety of different graph topologies $\mathscr{G}$: fully connected graphs, one dimensional chains, and finally Watts-Strogatz small world networks. Without loss of generality, I set $U_0 = 1$ as this simply corresponds to a rescaling of β.

Figure 8.2 (Left) shows the value of β_c, determined as the maximum of the variance of h, as a function of T for two different topologies (one dimensional and fully connected) and two different values of $\gamma \in \{1, 1.5\}$. The

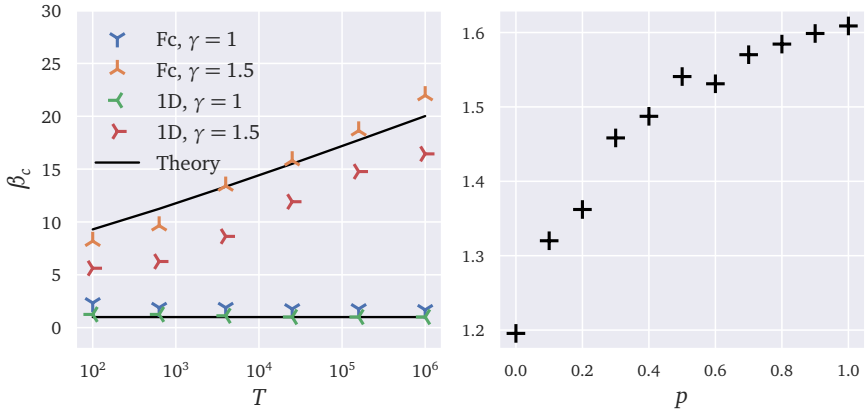


Figure 8.2: Left: critical β_c as a function of $\ln T$ for $\gamma = 1$ and 1.5, $N = 10^5$ and different topologies. (Fc stands for fully connected, while 1D is the one-dimensional chain). Black lines correspond to the prediction of mean-field theory. Right: dependence of the critical intensity of choice $\beta_c^\star$ on the parameter p of Watts-Strogatz networks, for $T = 5 \cdot 10^3$ and $N = 2 \cdot 10^3$.

results show excellent qualitative agreement with the theoretical prediction for the two topologies, in particular Eq. (8.10) in the $\gamma > 1$ case, although there is an overall factor needed to account for the one-dimensional data.

One can actually interpolate between the two situations by considering Watts-Strogatz small-world networks [269], with a rewiring parameter p such that $p = 0$ corresponds to one-dimension chains and $p = 1$ to the fully connected graph. Figure 8.2 (Right) shows the value of $\beta_c^\star$ as a function of the rewiring parameter p of interpolating between between one-dimensional chains for $p = 0$ and the fully connected graph for $p = 1$. The parameter p therefore allows to interpolate between a situation where one may only do local jumps to a situation where one can go anywhere. As expected, $\beta_c^\star$ increases with p, as it is easier to get trapped in less connected graphs, where the same choice is revisited more often.

I now take to studying more carefully the behaviour of the order parameter h close to the transition point $\beta_c^\star$, when $\gamma = 1$, both for one dimensional chains and for the fully connected graph. I choose $N = 10^5$ henceforth, such that finite size effects are negligible in the range of T is explored. Figure 8.3 suggests that as $T \to \infty$, $\langle h \rangle(\beta)$ appears to slowly converge to a step function that is zero for $\beta < \beta_c^\star$ and unity when $\beta > \beta_c^\star$, at least in the fully connected case where the speed of convergence is found to be $\sim T^{-1/3}$. In the one dimensional case, one cannot exclude with the available data that this limiting function remains continuous when $T \to \infty$.

8.5 AGING

I will now be a little more specific about the meaning of self-trapping for finite T when $\beta > \beta_c^\star$. The correct statement is that the system *ages*, in

the following sense [59, 188]: assume that the agent's choice at time T is a certain x_i and ask: What is the probability $\mathscr{P}(t,T)$ that the agent has never changed his/her mind between T and a later time $T+t$? In the free exploration phase $\beta < \beta_c^\star$, this probability is, for large T, independent of T: the process is time-translation invariant. In the trapped phase $\beta > \beta_c^\star$, $\mathscr{P}(t,T)$ can be estimated by appropriately generalizing Eq. (8.7). The result takes the following aging form (see Fig 8.4):

$$\mathscr{P}(t,T) \approx \exp\left(\frac{1}{a(T+t)^a} - \frac{1}{aT^a}\right), \quad a = \frac{\beta}{\beta_c^\star} - 1 > 0. \tag{8.11}$$

Note that in the regime $t \ll T$, $\mathscr{P}(t,T)$ is a function of t/T^{1+a} (precisely $1-\mathscr{P}(t,T) \approx t/T^{1+a}$, see dashed line in Fig. 8.4), a regime called *super-aging* [61] since the effective time for changing one's mind grows as $T^{\beta/\beta_c^\star}$, i. e. faster than the age T itself. This is quite interesting since we are not aware of simple models leading to such a super-linear aging behaviour. Hence, memory effects of the type discussed here might very well be an interesting lead to interpret experiments that show such a super-aging behaviour, such as those reported in [87].

Right at the transition point $\beta = \beta_c^\star$, one finds simple aging, i. e. a scaling function of t/T:

$$\mathscr{P}(t,T) \approx \frac{1}{1+\frac{t}{T}}, \qquad \beta = \beta_c^\star. \tag{8.12}$$

When the kernel has a finite norm and leads to a crossover rather than a true transition, aging will take place whenever $T \ll e^{\beta\|\phi\|}$ but revert to a normal time translation dynamics when $T \gg e^{\beta\|\phi\|}$ (see [188] for a similar situation). When $\gamma < 1$, on the contrary, relaxation is quasi-frozen for large T, in the sense that $\mathscr{P}(t,T) \approx 1 - t\exp\left(-\beta U_0 T^{1-\gamma}\right)$ when $t \ll T^\gamma$.

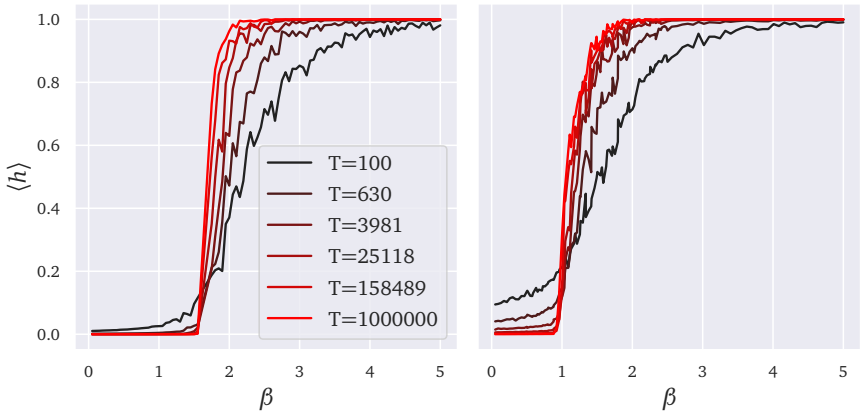


Figure 8.3: Order parameter h as a function β for $N = 10^5$, $\gamma = 1$ and different values of T. Left: Fully connected graph, for which a step function is approached as $T^{-1/3}$. Right: One-dimensional chain geometry, for which one cannot exclude that h remains a continuous function of β when $T \to \infty$.

Asymptotics of $\sum_{t=0}^{T}\phi(t)$	$\lvert\phi\rvert<\infty$	$\ln(T)$	$T^{1-\gamma}$, $\gamma<1$
Asymptotics of $P_{>}(\tau)$	$e^{-\lambda\tau}$ with $\lambda=\ln\left[1+e^{-\beta\lvert\phi\rvert}\right]$	• $\beta>\beta_c^\star$ (trapped regime): $P_\infty>0$ • $\beta=\beta_c^\star$ (critical regime): τ^{-1} • $\beta<\beta_c^\star$ (ergodic regime): $\exp\left(-\tau^b/b\right)$, with $b=1-\beta/\beta_c^\star$	$P_\infty>0$ (trapped)
Aging $\mathscr{P}(t,T)$	Trapping & aging for $T\ll e^{\beta\lvert\phi\rvert}$	• $\beta>\beta_c^\star$ super-aging $t\sim T^{\beta/\beta_c^\star}$ • $\beta=\beta_c^\star$ normal-aging $t\sim T$ • $\beta<\beta_c^\star$ equivalent to $\lvert\phi\rvert<\infty$	Quasi-frozen relaxation

Table 8.1: Summary of the different dynamical regimes.

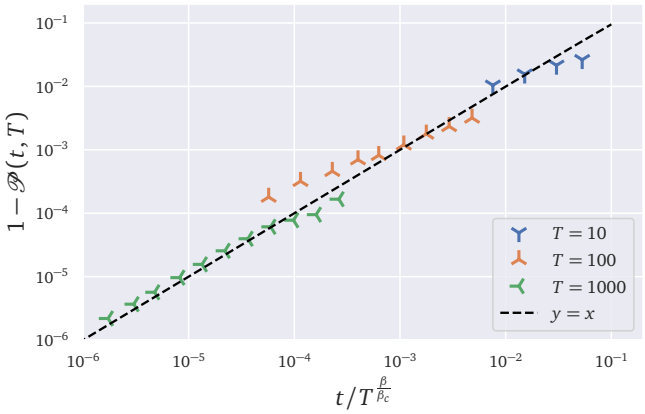


Figure 8.4: Aging function $1-\mathscr{P}(t,T)$ as a function of $t/T^{\beta/\beta_c^\star}$ in a fully connected graph with $N=10000$, with $\beta_c^\star \approx 1.5$ and $\beta = 3.5$. Eq. (8.11) predicts that $\mathscr{P}(t,T) \approx 1-t/T^{\beta/\beta_c^\star}$ when $t \ll T$ (dashed line).

8.6 CONCLUSION

Although quite simple, this model shows that non-trivial choice distortion effects can emerge through memory or self-reinforcing mechanisms. The main result is that the addition of memory effects can hinder the full exploration of all choices by the agent, and it may even cause him/her to leave a substantial number of possible options totally unexplored. The emergence of aging properties also shows that including memory effects in agents' preferences can lead to non-ergodic dynamics, when ergodicity is a crucial assumption to many models in economics.

Several extensions can be thought of, and would be a sensible way to incorporate more realism into the model. As mentioned, this model explores the case where the objective utility landscape $U_0(x)$ is totally flat, in a way to highlight the effects induced by memory alone. Reintroducing some heterogeneities in $U_0(x)$ would allow one to study the competition between "landscape trapping" and "memory trapping", with possibly interesting transitions between the two.

Another direction is to introduce many agents with interactions between them, meaning that the subjective utility may also depend on what others are doing. Here again, one expects some interesting competition between herding induced condensation of choices and memory effects. In particular, the combined effect of imitation of the past and imitation of peers may generate *collective self-fulfilling prophecies*.

Another direction one could explore is when the graph $\mathscr{G}$ defining the topology of the space of choices is itself time dependent – see [95] for a step in this direction. For example, the neighbourhood of each choice could be itself affected by past choices, or some new choices, not present initially, could present themselves later in time (for example, the opening of a new restaurant).

In all these cases, the basic question is whether memory effects, habit formation, or herding completely distort the choices dictated by their objective utilities or not. Such distortions may have very significant economic consequences at the macro level.

From a purely theoretical point of view, revisiting reinforcement models considered in the literature [209] with a power-law decaying memory kernel could lead to new interesting transitions of the type discussed above. In particular, the super-aging behaviour reported in the trapped phase might have applications much beyond the present setting.

CONCLUSION

CONCLUSION

In this thesis I have examined different possible explanations for the non-gaussian nature of macroeconomic fluctuations. Jumping back and forth between analytical models, numerical simulations and empirical data, I began by studying the origins and consequences of power-law distributions in socio-economic data and provided explanations for their pervasiveness and insights to some of the dynamical processes behind them. This also provided me with most of the intuition and technical skills required for the stochastic modelling that followed. I hope that the reader joined me in my amazement upon the discovery of these statistical laws that seem to be everywhere, and that remain an active subject of research as new mechanisms, in particular for the Zipf law, seem to pop up every now and then.

I then took to examining a network model that takes into account the intricacy of supply chain relations in a way that is strongly reminiscent of how different species in an ecosystem depend on one another, and that also highlights how a large economy may become intrinsically unstable through *self-organized criticality*. Although the version presented here is chiefly concerned with the static properties of a network economy, it laid the ground for further theoretical and empirical exploration. An ongoing work done within the *Econophysix* research chair explores a dynamical extension of this model and uncovers a plethora of regimes, including chaotic dynamics and other scenarios far removed from what is commonly considered in economic modelling. This should be the object of an article shortly after the publication of this thesis.

Following the analysis of this network model, I conducted an empirical study of the statistics of firm growth rates, the goal of which is ultimately to provide a framework with which to study production networks empirically. A remaining open question is how to weld these two approaches together, namely the empirical study of the firms that make up a country and the theoretical exploration of their possible trajectories given very general assumptions on how they function. My sincere hope is that in the long run this leads to an ability to assess the current status of the economy, to forecast its possible evolution and to provide tools and ideas to policy makers. The situation during the Covid crisis shows a stark contrast between our digital post-industrial society, the amount of data it generates and our current incapacity to have a clear image of its economic consequences and of how to mitigate them; the frustration that arises from this and other situations is a natural source of motivation, at least to me, to pursue this research program.

Finally, the last part of my thesis dealt with the exploration of models more steeped in the general subject of complexity, where I attempted to

study the consequences of including herding behaviour, ergodicity breaking and path dependence in simple agent-based models. This of course offers multiple research possibilities, and I believe that any should be pursued both for the intellectual satisfaction they may provide and as a mental exercise to build intuition in how dynamics defined at the scale of the individual translate to the scale of a collective.

This, I think, is part of the more general exercise of trying to make sense of the world that surrounds us: as individuals, our intuition is formed at the "microscopic" level at which we make our own decisions, and the often surprising things we observe at the level of the society happen at a scale to which we are not accustomed. To borrow from Schelling, it is a matter of making sense of *Macrobehaviour* from *Micromotives* [232] by doing so in the other direction, by playing God and creating a small world of agents to see what it becomes.

APPENDIX

A

SOME RESULTS IN PROBABILITY THEORY

A.1 A SKETCH OF THE PROOF FOR THE GENERALIZED CLT

For this section, I take a random variable with $\mathbb{E}[X] = 0$, when it is defined, without loss of generality. The starting point to understand the generalized CLT is to see that the distribution of the sum of two variables $S_2 = X_1 + X_2$ is given by the convolution of their densities. Call $p_2(s)$ the density of the sum, and $p(x)$ the density of the two variables X_1, X_2. One has clearly

$$p_2(s) = \int \mathrm{d}x_1 \mathrm{d}x_2 \; p(x_1)p(x_2)\delta(s-(x_1+x_2)) = \int \mathrm{d}x_1 \; p(x_1)p(s-x_1) \\ = (p)^{\star 2}(s), \tag{A.1}$$

where $(p)^{\star N}$ means the density p convoluted N times.

If we consider now the sum S_N of N iid copies of this variable

$$S_N = \sum_{i=1}^{N} X_i, \tag{A.2}$$

we can define its characteristic function $\phi_N(t) = \mathbb{E}\left[e^{itS_N}\right]$. Using the definition of the characteristic function $\phi_X(t) = \mathbb{E}\left[e^{itX}\right]$ and the fact that the Fourier transform maps convolutions into simple products, we see that

$$\phi_N(t) = \left[\phi_X(t)\right]^N. \tag{A.3}$$

When the moment $\sigma^2 = \mathbb{E}\left[X^2\right]$ is defined, we can expand the exponential for small t as

$$\phi_X(t) = \mathbb{E}\left[e^{itX}\right] \approx 1 - \frac{t^2\sigma^2}{2}, \tag{A.4}$$

and notice then that the characteristic function of the variable $S_N/\sqrt{N}$ is $\phi_N(t/\sqrt{N})$, and therefore given by

$$\phi_N\left(\frac{t}{\sqrt{N}}\right) \approx \left(1 - \frac{\sigma^2 t^2}{2N}\right)^N \approx e^{-\frac{\sigma^2 t^2}{2}}, \tag{A.5}$$

which is the characteristic function of a gaussian distribution. This shows then that $S_N/\sqrt{N}$ converges in law to a gaussian distribution, as claimed in Eq. (2.16).

When the density $p(x)$ decays as a power-law with $\mu \leq 2$ the second moment is no longer defined. To have a similar argument, one needs to compute the characteristic function $\phi_X(t)$ of X, or equivalently the Fourier transform of p, for small t. One can then show that in the case where X has symmetric positive and negative tails (and has 0 mean in the case $1 < \mu \leq 2$), with $p(x) \underset{x\to\pm\infty}{\approx} A|x|^{-1-\mu}$ its characteristic function for small t reads

$$\phi_X(t) \approx 1 - C|t|^{\mu}, \tag{A.6}$$

with $C \propto A$, so that the characteristic function for $S_N / N^{1/\mu}$ becomes

$$\phi_S\left(\frac{t}{N^{1/\mu}}\right) \approx \left(1 - \frac{C|t|^{\mu}}{N}\right)^N \approx \exp(-C|t|^{\mu}), \tag{A.7}$$

which can then be shown to be the Fourier transform of a power-law function behaving as $x^{-1-\mu}$ and thus preserving the power-law structure.

For large N, one can then write the generalized central limit theorem, stating that

$$S_N \approx \begin{cases} N\mathbb{E}[X] + (AN)^{1/2}\,\xi_G & \text{for} \quad \mu > 2 \\ N\mathbb{E}[X] + (AN)^{1/\mu}\,\xi_L & \text{for} \quad 1 < \mu \leq 2\,, \\ (AN)^{1/\mu}\,\xi_L & \text{for} \quad \mu < 1 \end{cases} \tag{A.8}$$

where ξ_L is a Lévy-stable distribution with tail parameter μ.

In general, the generalized CLT can be seen as the definition of functional fixed points for the convolution operation. For probability distributions decaying faster than $1/x^3$ this fixed point is necessarily gaussian, while distributions with a slower decay flow to fixed points defined by the Lévy-stable distributions.

A.2 SUM OF k LAPLACE VARIABLES

Consider the random variable $Y_k = \sum_{i=1}^{k} X_i / \sqrt{2k}$ where the X_i variables have a Laplace distribution with density

$$p(x) = \frac{1}{2}\exp(-|x|), \tag{A.9}$$

so that $\mathbb{E}[Y_k] = \sqrt{k}\mathbb{E}[X_i] = 0$ and $\mathbb{E}\left[Y_k^2\right] = \frac{\mathbb{E}\left[X_i^2\right]}{2} = 1$.

The characteristic function of the X_i variables is naturally given by

$$\phi_X(t) = \frac{1}{2}\int \mathrm{d}x\; e^{\mathrm{i}tx - |x|} = \frac{1}{1+t^2}, \tag{A.10}$$

and therefore the characteristic function of Y_k reads

$$\phi_{Y,k}(t) = \frac{2^k k^k}{(2k+t^2)^k}. \tag{A.11}$$

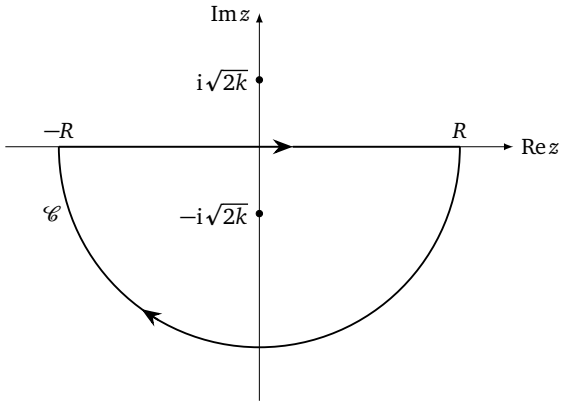


Figure A.1: Example of the contour taken in the case $y > 0$.

A.2.1 *Computing the density for Y_k*

To get the density for Y_k, $g_k(y)$ I next compute the inverse Fourier transform of its characteristic function, i. e. :

$$\begin{aligned} g_k(y) &= \frac{2^k k^k}{2\pi} \int \mathrm{d}t \; \frac{e^{-\mathrm{i}ty}}{(2k+t^2)^k} \\ &= \frac{2^k k^k}{2\pi} \int_{\mathscr{C}} \mathrm{d}z \; f(z) \end{aligned} \tag{A.12}$$

with $f(z) = \frac{e^{-\mathrm{i}zy}}{(2k+z^2)^k} = \frac{e^{-\mathrm{i}zy}}{\left(z-\mathrm{i}\sqrt{2k}\right)^k \left(z+\mathrm{i}\sqrt{2k}\right)^k}$, with a contour $\mathscr{C}$ running from $-R$ to R and then around the half-circle centred at the origin and of radius R with $R > \sqrt{2k}$. The contour is chosen so that it encloses $-\mathrm{i}\sqrt{2k}$ if $y > 0$ and $\mathrm{i}\sqrt{2k}$ if $y < 0$, and then the limit $R \to \infty$ is taken. See Figure A.1 for an illustration for $y > 0$.

In this case, the function $f(z)$ has a pole of order k at $z = -\mathrm{i}$, of residue

$$\mathrm{Res}_{-\mathrm{i}\sqrt{2k}}(f) = \frac{1}{(k-1)!} \lim_{z \to -\mathrm{i}\sqrt{2k}} \frac{\partial^{(k-1)}}{\partial z^{(k-1)}} \left(\frac{e^{-\mathrm{i}zy}}{\left(z-\mathrm{i}\sqrt{2k}\right)^k} \right). \tag{A.13}$$

To compute the derivative I now apply the generalized Leibniz' rule

$$\frac{\partial^{(k-1)}}{\partial z^{(k-1)}} \left(\frac{e^{-\mathrm{i}zy}}{\left(z-\mathrm{i}\sqrt{2k}\right)^k} \right) = \sum_{l=0}^{k-1} \binom{k-1}{l} \frac{\partial^l}{\partial z^l} \left(e^{-\mathrm{i}zy}\right) \frac{\partial^{(k-1-l)}}{\partial z^{(k-1-l)}} \left(\left(z-\mathrm{i}\sqrt{2k}\right)^{-k} \right) \tag{A.14}$$

leading to:

$$\mathrm{Res}_{-\mathrm{i}}(f) = \mathrm{i} \frac{1}{(k-1)!} \sum_{l=0}^{k-1} \binom{k-1}{l} \left(2\sqrt{2k}\right)^{l+1-2k} \frac{(2k-2-l)!}{(k-1)!} y^l e^{-\sqrt{2k}y}.$$

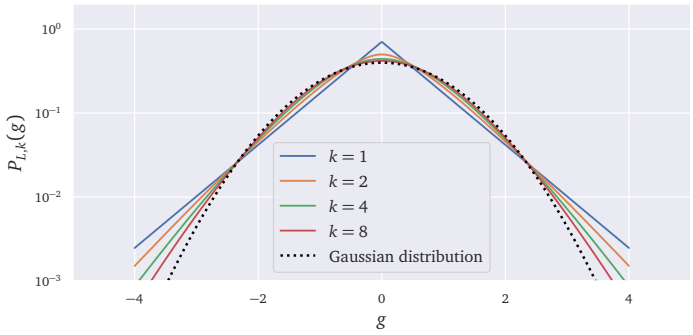


Figure A.2: Examples of the sum of k Laplace variables for $k = \{1, 2, 6, 8\}$ with a Gaussian for comparison. Note that the Laplace "cusp" is already gone at $k = 2$.

(A.15)

Plugging this into $g_k(y) = \frac{(2k)^k}{2\pi}\int_{\mathscr{C}} f = -\mathrm{i}(2k)^k \mathrm{Res}_{-\mathrm{i}}(f)$, and symmetrizing when $y < 0$

$$g_k(y) = \frac{2^{-2k}2\sqrt{2k}}{(k-1)!}\sum_{l=0}^{k-1}\binom{k-1}{l}2^l\frac{(2k-2-l)!}{(k-1)!}(\sqrt{2k}|y|)^l e^{-\sqrt{2k}|y|}. \tag{A.16}$$

The behaviour of this density is depicted on Figure A.2, where it is clear that the Laplace cusp is not stable upon aggregation of variables.

A.2.2 *Extension to sums of correlated Laplace variables*

Suppose now that we want to compute the distribution of the sum of two Laplace random variables X_1 and X_2, with correlation $\mathbb{E}(X_1X_2) = \rho$, and where both variables have the density given in Eq. (A.9).

To tackle this problem, we first prove a preliminary result, which is that the variables X can be represented as a superposition of gaussian random variables with variances distributed according to an exponential distribution, namely

$$X \overset{\text{law}}{=} \mathbb{E}_\sigma\left[Y_\sigma\right], \tag{A.17}$$

where the distribution of Y_σ and of σ[1] are given by

$$p_{Y_\sigma}(y) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-y^2/2\sigma^2}, \quad p_\sigma(\sigma) = \sigma e^{-\sigma^2/2}\mathbf{1}_{\sigma>0}. \tag{A.18}$$

1 Mind that σ is the square root of the variance, and that it is only its square that is distributed exponentially.

To prove this, compute the characteristic function of X as

$$\begin{aligned}\phi_X(t) &= \int_0^\infty \mathrm{d}\sigma\ \sigma e^{-\sigma^2/2}\left[\int \frac{\mathrm{d}y}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{y^2}{2\sigma^2}+ity\right)\right] \\ &= \int_0^\infty \mathrm{d}\sigma\ \sigma\exp\left(-\frac{\sigma^2}{2}\left(1+t^2\right)\right) \\ &= \frac{1}{1+t^2},\end{aligned} \tag{A.19}$$

which proves the previous assertion.

I now generalize to try to find the sum of k correlated Laplace variables. To do this, consider the sum $\sum_{l=1}^k \sigma_l Y_l$ where Y_0 is gaussian and

$$Y_l = \rho Y_{l-1} + \sqrt{1-\rho^2}\eta_l, \tag{A.20}$$

with η_l a gaussian white noise of unit variance and σ_l distributed according to the density in Eq. (A.18). One sees that in this case the variables $\sigma_l Y_l$ are Laplace-distributed, with $\mathrm{Cor}(\sigma_l Y_l, \sigma_{l+1}Y_{l+1}) = 2\rho$.

For $k=2$ the sum is

$$S_2 = \sigma_1 Y_1 + \sigma_2 Y_2 = \sigma_1 Y_1 + \rho\sigma_2 Y_1 + \sigma_2\sqrt{1-\rho^2}\eta_2, \tag{A.21}$$

which, owing to the gaussian nature of Y_1, Y_2 and η_2 can be written as

$$S_2 = \sqrt{\left(\sigma_1+\rho\sigma_2\right)^2 + \sigma_2^2(1-\rho^2)}\xi, \tag{A.22}$$

with ξ gaussian with unit variance.

This allows to compute the characteristic function of S_2 given (σ_1, σ_2) as

$$\begin{aligned}\phi_2(t,\sigma_1,\sigma_2) =& \mathbb{E}\left[\exp\left(\mathrm{i}tS_2\right)|\sigma_1,\sigma_2\right] \\ =& \exp\left(-\frac{\left(\sigma_1+\rho\sigma_2\right)^2+\sigma_2^2\left(1-\rho^2\right)}{2}t^2\right),\end{aligned} \tag{A.23}$$

which I then average to obtain the characteristic function of S_2, as

$$\begin{aligned}\phi_2(t) &= \int_0^\infty \mathrm{d}\sigma_1\ \sigma_1 e^{-\frac{\sigma_1^2}{2}}\int_0^\infty \mathrm{d}\sigma_2\ \sigma_2 e^{-\frac{\sigma_2^2}{2}}\phi_2(t,\sigma_1,\sigma_2) \\ &= \int_0^\infty \mathrm{d}\sigma_1 \int_0^\infty \mathrm{d}\sigma_2\ \ \sigma_1\sigma_2\exp\left(-\frac{1}{2}\langle\sigma|A|\sigma\rangle\right),\end{aligned} \tag{A.24}$$

where $A = \begin{pmatrix} 1+t^2 & t^2\rho \\ t^2\rho & 1+t^2\end{pmatrix}$, with eigenvectors $|u\rangle = \frac{1}{\sqrt{2}}(1,1)$ and $|v\rangle = \frac{1}{\sqrt{2}}(1,-1)$ and eigenvalues $\lambda_1 = 1+t^2(1-\rho)$ and $\lambda_2 = 1+t^2(1+\rho)$, where the dependence in t of λ_1, λ_2 is implicit.

I therefore substitute $u = \frac{\sqrt{2}}{2}(\sigma_1 + \sigma_2)$ and $v = \frac{\sqrt{2}}{2}(\sigma_1 - \sigma_2)$ in Eq. (A.24), to obtain

$$\begin{aligned}\phi_2(t) =& \frac{1}{2}\int \mathrm{d}v \int \mathrm{d}u \, \left(u^2 - v^2\right) \exp\left(-\lambda_1 \frac{u^2}{2} - \lambda_2 \frac{v^2}{2}\right) \mathbf{1}_{u+v>0} \mathbf{1}_{u-v>0} \\ =& \frac{1}{2}\int \mathrm{d}v \int \mathrm{d}u \, u^2 \exp\left(-\lambda_1 \frac{u^2}{2} - \lambda_2 \frac{v^2}{2}\right) \mathbf{1}_{u+v>0} \mathbf{1}_{u-v>0} \\ &+ \frac{1}{2}\int \mathrm{d}v \int \mathrm{d}u \, v^2 \exp\left(-\lambda_1 \frac{u^2}{2} - \lambda_2 \frac{v^2}{2}\right) \mathbf{1}_{u+v>0} \mathbf{1}_{v-u>0}.\end{aligned} \tag{A.25}$$

Computing now these terms explicitly,

$$\begin{aligned}&\int \mathrm{d}u \, e^{-\lambda_1 \frac{u^2}{2}} u^2 \int \mathrm{d}v \, e^{-\lambda_2 \frac{v^2}{2}} \mathbf{1}_{u+v>0} \mathbf{1}_{u-v>0} \\ =& 2\sqrt{\frac{\pi}{2\lambda_2}} \int_0^\infty \mathrm{d}u \, e^{-\lambda_1 \frac{u^2}{2}} u^2 \mathrm{erfc}\left(\sqrt{\frac{\lambda_2}{2}} u\right) \\ =& 4\sqrt{\frac{\pi}{\lambda_2}} \frac{1}{\lambda_1^{3/2}} \int_0^\infty \mathrm{d}w \, e^{-w^2} w^2 \mathrm{erfc}\left(\sqrt{\frac{\lambda_2}{\lambda_1}} w\right),\end{aligned} \tag{A.26}$$

with the last integral giving (see e. g. [200])

$$\int_0^\infty \mathrm{d}w \, e^{-w^2} w^2 \mathrm{erfc}\left(\sqrt{\frac{\lambda_2}{\lambda_1}} w\right) = \frac{1}{2\sqrt{\pi}} \left(\arctan\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right) - \frac{\lambda_1}{\lambda_2 + \lambda_1} \sqrt{\frac{\lambda_2}{\lambda_1}}\right), \tag{A.27}$$

so that Eq. (A.26) reads

$$\begin{aligned}&\int \mathrm{d}u \, e^{-\lambda_1 \frac{u^2}{2}} u^2 \int \mathrm{d}v \, e^{-\lambda_2 \frac{v^2}{2}} \mathbf{1}_{u+v>0} \mathbf{1}_{u-v>0} \\ =& \frac{2}{\lambda_1^{3/2} \lambda_2^{1/2}} \left(\arctan\left(\sqrt{\frac{\lambda_1}{\lambda_2}}\right) - \frac{\lambda_1}{\lambda_2 + \lambda_1} \sqrt{\frac{\lambda_2}{\lambda_1}}\right).\end{aligned} \tag{A.28}$$

Gathering everything, the result is

$$\begin{aligned}\phi_2(t) =& \frac{\arctan\left(\sqrt{\lambda_1/\lambda_2}\right)}{\lambda_1^{3/2}\lambda_2^{1/2}} + \frac{\arctan\left(\sqrt{\lambda_2/\lambda_1}\right)}{\lambda_2^{3/2}\lambda_1^{1/2}} \\ &- \frac{1}{\sqrt{\lambda_1\lambda_2}\,(\lambda_1 + \lambda_2)} \left(\sqrt{\frac{\lambda_2}{\lambda_1}} + \sqrt{\frac{\lambda_1}{\lambda_2}}\right).\end{aligned} \tag{A.29}$$

For $|\rho| < 1$ can nevertheless write $\lambda_1(t) = \left(1 + t^2\right)\left(1 - \frac{t^2}{1+t^2}\rho\right)$ and $\lambda_2(t) = \left(1 + t^2\right)\left(1 + \frac{t^2}{1+t^2}\rho\right)$, and so one sees that $\phi_2(t) \propto (1 + t^2)^{-2}$ quickly as t grows large. When $|\rho| = 1$ however, we have e.g. $\lambda_1 = 1$ and $\lambda_2 = 1 + 2t^2$, and as a consequence $\phi_2(t) \sim 1/t^3$ for large t. One

can look at the curvature of the distribution function $p_2(s)$ of the sum S_2 defined in Eq. (A.21), as

$$p_2''(0) = \int \mathrm{d}t \; t^2 \phi(t), \tag{A.30}$$

and see that it behaves as $(1-|\rho|)^{-3/2}$.

B

FOKKER-PLANCK AND LANGEVIN EQUATIONS

B.1 A QUICK INTRODUCTION TO ITÔ CALCULUS

In the modelling questions we have faced, one is often interested in the stochastic evolution of a quantity $x(t)$, an evolution that can be written in a discrete form as

$$x(t+\Delta t) = x(t) + a(x,t)\Delta t + b(x,t)\eta(t)\Delta t, \tag{B.1}$$

where $\eta(t)$ is a gaussian white noise, meaning $\mathbb{E}\left[\eta(t)\Delta t\right] = 0$ and $\mathbb{E}\left[\eta(t)\eta(t')\right] = \delta(t-t')$.

The function $a(x,t)$ dictates the deterministic part of the evolution of x, while $b(x,t)\eta(t)$ is the stochastic part. One is then tempted to look at the limit $\lim_{\Delta t\to 0} \frac{x(t+\Delta t)-x(t)}{\Delta t}$, and write it as

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a(x,t) + b(x,t)\eta(t). \tag{B.2}$$

But we should be careful. In Eq. (B.1) one can write the displacement $\Delta x = x(t+\Delta t) - x(t)$ between t and $t+\Delta t$, and check that

$$\mathbb{E}\left[\Delta x\right] = a(x,t)\Delta t, \tag{B.3}$$

giving a good limit to the average of Eq. (B.2), as $\mathbb{E}\left[\Delta x\right]/\Delta t$ is of order 1 as $\Delta t \to 0$.

Yet what about it's variance ? Typically, one should expect that $\mathbb{V}\left[\Delta x\right]/\Delta t$ remains of order 1 as $\Delta t \to 0$, so that Δx remains a random quantity in that limit. This implies that

$$\mathbb{E}\left[b(x,t)^2\eta(t)^2\Delta t^2\right] \propto \Delta t, \tag{B.4}$$

and therefore that $\eta(t) \propto 1/\sqrt{\Delta t}$. This is the main reason for the "weird" aspect of Itô calculus.

B.1.1 *Taylor expansions*

With this $\eta(t) \propto 1/\sqrt{\Delta t}$ scaling in mind, how should one understand the variation of a function $f(x(t))$ between t and $t+\Delta t$?

One is of course tempted to write

$$\Delta f \approx \frac{\mathrm{d}f}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t}\Delta t + \mathcal{O}(\Delta t^2), \tag{B.5}$$

but this is false, as the squared terms in the Taylor expansion would give a term in $\eta(t)^2 \propto \Delta t$. Instead, one should go to second order, as

$$\Delta f \approx \frac{\mathrm{d}f}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t}\Delta t + \frac{1}{2}\frac{\mathrm{d}f}{\mathrm{d}x}\left(\frac{\mathrm{d}x}{\mathrm{d}t}\Delta t\right)^2, \tag{B.6}$$

which can then be rewritten as

$$\Delta f = \left[-a(x,t)\frac{\mathrm{d}f}{\mathrm{d}x} + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}b(x,t)^2 + b(x,t)\frac{\mathrm{d}f}{\mathrm{d}x}\eta(t)\right]\Delta t, \tag{B.7}$$

which is the new stochastic differential equation for f.

B.1.2 *Deriving the Fokker-Planck equation*

How can one use the previous result to find how the probability distribution $p(x,t)$ describing the variable $x(t)$ evolves ? The Taylor expansion above is independent of the function f. We pick f to be any well behaved function with $\lim_{\pm\infty} f = 0$, and we write on the one hand

$$\begin{aligned}\mathbb{E}[\Delta f] &= \int \mathrm{d}x\ (p(x,t+\Delta t) - p(x,t))f(x) \\ &= \left[\int \mathrm{d}x\ \frac{\partial p(x,t)}{\partial t}f(x)\right]\Delta t,\end{aligned} \tag{B.8}$$

but also

$$\begin{aligned}\mathbb{E}[\Delta f] &= \mathbb{E}\left[a(x,t)\frac{\mathrm{d}f}{\mathrm{d}x} + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}b(x,t)^2 + b(x,t)\frac{\mathrm{d}f}{\mathrm{d}x}\eta(t)\right]\Delta t \\ &= \int \mathrm{d}x\ p(x,t)\left[a(x,t)\frac{\mathrm{d}f}{\mathrm{d}x} + \frac{1}{2}\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}b(x,t)^2\right]\Delta t \\ &= \int \mathrm{d}x\ f(x)\left(-\frac{\partial}{\partial x}[a(x,t)p(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b(x,t)^2 p(x,t)\right]\right)\Delta t,\end{aligned} \tag{B.9}$$

where on passing from the first to the second line we used that $\mathbb{E}[\eta(t)] = 0$, and where we integrated by parts to get to the last line.

Equalling both expressions we get

$$\begin{aligned}\int \mathrm{d}x\ \frac{\partial p(x,t)}{\partial t}f(x) = \int \mathrm{d}x\ \Bigg(&-\frac{\partial}{\partial x}[a(x,t)p(x,t)] \\ &+ \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b(x,t)^2 p(x,t)\right]\Bigg)f(x),\end{aligned} \tag{B.10}$$

which, being true for *any* function f, implies the Fokker-Planck equation

$$\frac{\partial p(x,t)}{\partial x} = -\frac{\partial}{\partial x}[a(x,t)p(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b(x,t)^2 p(x,t)\right]. \tag{B.11}$$

B.1.3 *Standard Langevin equations*

The Langevin equation was proposed in 1908 as a way to model brownian motion. An interesting aspect of it is that it gives a clear physical interpretation to some stochastic differential equations. If one imagines a particle in, for example, a colloidal suspension, then its equation of motion is given by

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -\gamma\frac{\mathrm{d}x}{\mathrm{d}t} + F(x) + \eta(t), \tag{B.12}$$

with m its mass, γ a friction coefficient, F a force field and $\eta(t)$ random fluctuations due to the environment (think of collisions with smaller particles in the suspension). In the Itô prescription, the correlation time of these random fluctuations is $\tau_C \ll \mathrm{d}t$, and therefore it can be modelled as gaussian noise with $\mathbb{E}[\eta(t)] = 0$ and $\mathbb{E}[\eta(t)\eta(t')] = 2T\delta(t-t')$, where T is the temperature. In the over-damped limit $\frac{m}{\gamma} \ll 1$ the $\ddot{x}$ term is negligible and the equation becomes:

$$\begin{aligned} \gamma\frac{\mathrm{d}x}{\mathrm{d}t} &= F(x) + \eta(t) \\ \frac{\mathrm{d}x}{\mathrm{d}t} &= -V'(x(t)) + \eta(t), \end{aligned} \tag{B.13}$$

where in the second line we've set $\gamma = 1$ (always possible with a suitable change of units) and supposed that $F(x) = -V'(x)$ derives from a potential V.

From the previous part, we see that the corresponding Fokker-Planck equation is given by

$$\frac{\partial p(x,t)}{\partial x} = \frac{\partial}{\partial x}\left[V'(x)p(x,t)\right] + T\frac{\partial^2 p(x,t)}{\partial x^2}, \tag{B.14}$$

and its stationary state can then be found by setting $\partial_t p(x,t) = 0$, as

$$p(x,t) = \frac{\exp\left(-\frac{V(x)}{T}\right)}{Z}, \tag{B.15}$$

with Z a normalizing constant, called the partition function in statistical mechanics. One therefore finds the Boltzmann-Gibbs distribution corresponding to the potential $V(x)$.

The advantage of the Langevin Eq. (B.13) is that it gives a stochastic process a clear interpretation: that of a particle doing a "stochastic gradient descent" (going towards a local minimum of V but with fluctuations).

B.2 TRANSITION MATRICES AND FOKKER-PLANCK OPERATORS

B.2.1 *The Fokker-Planck operator and transition matrices*

In all generality, one can write the master equation for a markovian evolution as

$$\frac{\partial p(x,t)}{\partial t} = \int \mathrm{d}y \ \left(W(y \to x)\, p(y,t) - W(x \to y) p(x,t)\right), \quad \text{(B.16)}$$

where W is the continuum equivalent of the transition matrix, i. e. $W(y \to x)$ is the probability of jumping from y to x during a time $\mathrm{d}t$. The intuition then is that for some processes the terms $W(y \to x)$ and $W(x \to y)$ are nearly zero for $|y - x| \gg 1$. We therefore denote

$$W_{-r}(x) = W(x \to x - r), \quad \text{(B.17)}$$

and so the new equation reads

$$\frac{\partial p(x,t)}{\partial t} = \int \mathrm{d}r \ \left(W_r(x-r) p(x-r,t) - W_r(s) p(x,t)\right). \quad \text{(B.18)}$$

We may now Taylor-expand the right-hand side in r, yielding

$$\frac{\partial p(x,t)}{\partial t} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial}{\partial x^n} \left[a_n(x) p(x,t)\right], \quad \text{(B.19)}$$

also known as the Kramers-Moyal development, and the a_n coefficients are

$$a_n(x) = \int \mathrm{d}r \ r^n W_{-r}(x), \quad \text{(B.20)}$$

the moments of the probability distribution describing the size r of a jump starting from x. When the process does not have any discontinuous jumps, viz. when the η variable in Eq. (B.1) has a finite second moment in the limit $\Delta t \to 0$, it can be shown that one need only keep the terms for $n = 1, 2$ [226].

This leads again to the Fokker-Planck Eq. (B.11), but importantly it shows that it can be written using an operator $\mathscr{A}$,

$$\frac{\partial p}{\partial t} = \mathscr{A} p, \quad \text{(B.21)}$$

which can in fact be thought of as the continuum equivalent of the transition matrix in standard Markov processes.

Indeed, in the case of discrete dynamics, one can write a markovian evolution as

$$\left|P(t+1)\right\rangle = \mathbf{T} \left|P(t)\right\rangle, \quad \text{(B.22)}$$

where $(|P\rangle)_i := \langle i|P|i|P\rangle$ is the probability associated with state i and $(\mathbf{T})\,ij$ is the probability of the transition $j \to i$, thereby defining the transition matrix.

Roughly, if one can diagonalize $\mathbf{T}$ as

$$\mathbf{T} = |u_1\rangle\langle v_1| + \sum_{\lambda\neq 1} \lambda\,|u_\lambda\rangle\langle v_\lambda|\,, \tag{B.23}$$

with $|v_\lambda\rangle$ and $|u_\lambda\rangle$ the left and right eigenvectors associated to λ. The fact that $\lambda = 1$ is a (maximal) eigenvalue of this matrix is ensured by the Gershgorin circle theorem and the Perron-Frobenius theorem.

Successively applying this matrix to $\big|P(0)\big\rangle$ immediately gives for large T that

$$\big|P(t)\big\rangle \approx |u_1\rangle + \lambda_{\max}^t\,|\delta P\rangle\,, \tag{B.24}$$

where $\lambda_{\max} = \max\{\lambda \in \mathrm{Sp}(\mathbf{T}), |\lambda| < 1\}$. Hence, the stationary distribution is given by the vector $|u_1\rangle$, and the convergence time is exponential, at a rate $\ln(\lambda_{\max}) < 0$.

The same can be said of Eq. (B.21), and one can look at the spectrum of the operator $\mathscr{A}$ by solving the eigenvalue problem

$$\mathscr{A}p_{\mathscr{E}} = \mathscr{E}p_{\mathscr{E}}, \tag{B.25}$$

under the condition that $\int p_{\mathscr{E}} = 1$. This can sometimes be done directly, but the difficulty is easily understood when considering the diagonalization in Eq. (B.23). Indeed, as the transition matrix $\mathbf{T}$ is often not symmetric, it is also quite common that the operator $\mathscr{A}$ is not *self-adjoint*. Just as we can define the transpose ${}^t\mathbf{T}$ of a matrix by stating that it satisfies

$$\langle u|\mathbf{T}v\rangle = \big\langle {}^t\mathbf{T}u\big|v\big\rangle, \tag{B.26}$$

we can define the adjoint $\mathscr{A}^*$ of $\mathscr{A}$ by stating that for any two functions f, g,

$$\int g\,(\mathscr{A}f) = \int (\mathscr{A}^* g)\,f\,. \tag{B.27}$$

For example, in the case of simple diffusion with drift,

$$\mathscr{A}p(x,t) = -\mu\frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 p(x,t)}{\partial x^2}, \tag{B.28}$$

the adjoint can be found using Eq. (B.27) by integrating by parts, yielding

$$\mathscr{A}^* p(x,t) = \mu\frac{\partial p(x,t)}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 p(x,t)}{\partial x^2}. \tag{B.29}$$

B.2.2 *Transforming into a Schrödinger's equation*

Self-adjoint (also called Hermitian) operators, essentially those for which $\mathscr{A}^* = \mathscr{A}$ are often much nicer to work with. They are the equivalent of symmetric matrices, and possess much of their properties. In particular, under additional hypotheses, their eigenvalues and eigenvectors have interesting properties. In the cases that interest us, the set of numbers $\mathscr{E}$ satisfying Eq. (B.25) can be continuous or discrete, but taking for example the case where they are discrete, we could solve the Fokker-Planck equation as

$$p(x,t) = \sum_n \alpha_n f_n(x) e^{-\mathscr{E}_n t}, \tag{B.30}$$

where $\mathscr{E}_n$ are (minus) the eigenvalues and f_n the corresponding eigenvectors, or rather eigenfunctions. The eigenvalue $\mathscr{E}_0 = 0$ and its eigenfunction f_0 therefore correspond to the stationary state.

In particular, a well known class of self-adjoint operators is the family of quantum-mechanical hamiltonians, of the form

$$\mathbf{H}\Psi(x,t) = -\frac{\partial^2 \Psi(x,t)}{\partial x^2} + V(x)\Psi(x,t), \tag{B.31}$$

which often have solutions that can be found in the literature.

A common trick to solve Fokker-Planck equations therefore is to transform them into Schrödinger's equations. We start from a generic Fokker-Planck equation such as the one defined in Eq. (B.11), but with constant $b(x,t) = 2$ and time-independent drift, which we represent with the derivative of some function A, $a(y,t) = -A'(y)$. The resulting Fokker-Planck equation reads

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}\left[A'(x)p(x,t)\right] + \frac{\partial^2 p(x,t)}{\partial x^2} \tag{B.32}$$

and has a stationary solution that can be written as a Boltzmann distribution $p_0(y) = e^{-A(y)}/Z$, where Z is a constant ensuring normalisation.

We next introduce a function Ψ verifying $p(x,t) = e^{-A(x)/2}/\sqrt{Z}\Psi(x,t)$. We can compute derivatives to find

$$\begin{aligned}
\partial_x\left[A'(x)p(x,t)\right] = e^{-A(x)/2}/\sqrt{Z}\Bigg[&\left(A''(x) - \frac{A'(x)^2}{2}\right)\Psi(x,t) \\
&+ A'(x)\partial_x\Psi(x,t)\Bigg] \\
\partial_{xx}p(x,t) = e^{-A(x)/2}/\sqrt{Z}\Bigg[&-\frac{1}{2}\left(A''(x) - \frac{A'(x)^2}{2}\right)\Psi(x,t) \\
&- A'(x)\partial_x\Psi(x,t) + \partial_{xx}\Psi(x,t)\Bigg].
\end{aligned}$$

(B.33)

Adding these terms and simplifying, we find the following Schrödinger's equation for Ψ:

$$-\partial_t \Psi(y,t) = \mathbf{H}\Psi, \tag{B.34}$$

where the Hamiltonian is here defined as

$$\mathbf{H} = -\partial_{xx} + V(x), \quad V(x) = -\frac{1}{2}\left(A''(x) - \frac{A'(x)^2}{2}\right). \tag{B.35}$$

This then allows the transformation of a Fokker-Planck problem into a quantum-mechanical problem, as one can now solve the equation

$$\mathbf{H}\Psi = \mathscr{E}\Psi, \tag{B.36}$$

with $\mathscr{E} = 0$ allowing to find the stationary state, and the smallest eigenvalue $\mathscr{E}_1$ giving the convergence time, as one should expect that the corrective terms in Eq. (B.30) decay as $e^{-\mathscr{E}_1 t}$ at long times.

C

SOME BASICS ON RANDOM-MATRIX THEORY

The following Appendix draws heavily from Livan, Novaes and Vivo's wonderful review [168], from the unpublished textbook by Bouchaud and Potters [222] and from unpublished work done in conjunction with (and also mostly by) Luís Carlos García del Molino while he was finishing his postdoc at CFM. More mathematically inclined readers can consult Terence Tao's course [259].

The first section is an introduction to the basic tools of random matrix, seeking mostly to give an intuition of their use and of the main results they can give. It will therefore be restricted to the analysis of real symmetric random matrices, with a special emphasis on the Gaussian Orthogonal Ensemble (GOE). I will begin by introducing these matrices and their probability measures before talking about their eigenvalue density and the Dyson brownian motion as a way to understand the behaviour of the eigenvalues and eigenvectors of GOE matrices. I will then give a quick overview on localization and the mechanism behind it.

C.1 REAL SYMMETRIC RANDOM MATRICES

A matrix is, simply put, a 2-dimensional array of numbers, and a random matrix is simply a matrix where these numbers are drawn from a certain distribution, or are functions of other random numbers. A natural case, with direct applications in risk management, consists of looking at matrices of the form

$$\mathbf{M} = \frac{1}{T}\xi\xi^t, \tag{C.1}$$

where ξ is an $N \times T$ matrix consisting of N time-series of length T. When each time-series consists of gaussian white noise $\mathbf{M}$ it is possible to compute its spectrum, with the first given by the Marčenko-Pastur distribution [178]. This goes beyond the scope of the present appendix, but the interested reader can look into Ref. [222].

The simplest setting consists instead of having independent entries, but imposing symmetric coefficients to avoid dealing with complex eigenvalues. Not imposing symmetry would of course allow for a more realistic setting to describe, e. g. , matrices describing production networks or more general interactions, but the main intuition we wish to convey in this section remains unchanged. In this case, we consider the coefficients of the $N \times N$ matrix $\mathbf{M}$ to be drawn from a certain distribution, denoted $p(M_{ij})$, while also imposing $M_{ij} = M_{ji}$.

We can then introduce the Gaussian Orthogonal Ensemble, or GOE, where the coefficients are Gaussian and satisfy

$$\mathbb{E}\left[M_{ij}\right]=0, \quad \mathbb{E}\left[M_{ij}M_{kl}\right]=\frac{\sigma^2}{2N}\left(\delta_{il}\delta_{jk}+\delta_{ik}\delta_{jl}\right), \tag{C.2}$$

which means that the distribution $p(M_{ij})$ is a centred gaussian of variance $\frac{\sigma^2}{N}$ for diagonal coefficients M_{ii} and $\frac{\sigma^2}{2N}$ on the off-diagonal with $i \neq j$, with terms M_{ij} and M_{kl} decorrelated if $(i,j) \neq (k,l)$ or (l,k). Naturally, $M_{ij} = M_{ji}$ here.

In this case, it is easy to check that the probability distribution associated to the "whole" matrix $\mathbf{M}$, defined as the product over the probabilities of each of its components, is

$$p(\mathbf{M}) \propto \exp\left(-\frac{N}{2\sigma^2}\operatorname{Tr}\mathbf{M}^2\right). \tag{C.3}$$

This is why it's called the Gaussian *Orthogonal* Ensemble, since for any $N \times N$ matrix verifying $\mathbf{O}\mathbf{O}^t = \mathbf{1}$, namely any orthogonal matrix, it is immediate that $\mathbf{M}$ and $\mathbf{M}' = \mathbf{O}\mathbf{M}\mathbf{O}^t$ have the same probability.

Another ensemble of interest is that of sparse matrices, i. e. matrices for which the probability of their components is given by

$$p(M_{ij}) = \left(1-\frac{c}{N}\right)\delta\left(M_{ij}\right)+\frac{c}{N}\delta\left(M_{ij}-J\right), \tag{C.4}$$

with $M_{ii} = 0$ and $M_{ij} = M_{ji}$. This can be seen as the adjacency matrix of a random Erdös-Renyi graph, where each link has a weight J and each node is linked to c nodes on average.

C.1.1 *The resolvent matrix and Stieltjes transform*

Having defined the distributions of the $N(N+1)/2$ iid components of a random symmetric matrix, how should we go about studying the distribution of their N real eigenvalues? For each matrix realization from these ensembles, we can introduce the following probability measure

$$\rho(\lambda) = \frac{1}{N}\sum_{i=1}^{N}\delta\left(\lambda-\lambda_i\right), \tag{C.5}$$

where the λ_i are the eigenvalues corresponding to a specific realization of $\mathbf{M}$. For the matrices previously defined, ρ converges to a certain distribution, but we keep this discrete definition to aid in the understanding of what follows.

Instead of "counting" the eigenvalues equal to λ with the Dirac delta distribution, we can instead think of using a distribution δ_ε, counting the number of eigenvalues of $\mathbf{M}$ that fall within a value $\approx \varepsilon$ of λ, as one would do for example when using a Gaussian kernel density estimator of

resolution $\propto \varepsilon$. We can instead imagine using a Lorentzian of width ε, for reasons that will become clearer later, as

$$\rho_\varepsilon(\lambda) = \frac{1}{\pi\varepsilon}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{1+\left(\frac{\lambda-\lambda_i}{\varepsilon}\right)^2}, \tag{C.6}$$

and note that $\rho_\varepsilon \underset{\varepsilon\to 0}{\longrightarrow} \rho$, because a Lorentzian of width ε converges to the Dirac delta distribution as $\varepsilon \to 0$.

Noting however that for real z,

$$\frac{1}{z-\mathrm{i}\varepsilon} = \frac{z^2}{z^2+\varepsilon^2} + \frac{\mathrm{i}}{\varepsilon}\frac{1}{1+z^2/\varepsilon^2}, \tag{C.7}$$

which in fact explains that

$$\frac{1}{z-\mathrm{i}\varepsilon} \underset{\varepsilon\to 0}{\longrightarrow} \text{p.v.}\left(\frac{1}{z}\right) + \mathrm{i}\pi\delta(z) \tag{C.8}$$

where p.v. denotes the Cauchy principal value. This in turn motivates rewriting Eq. (C.6) as

$$\rho_\varepsilon(\lambda) = \frac{1}{\pi}\,\mathrm{Im}\left[\sum_{i=1}^{N}(\lambda-\mathrm{i}\varepsilon-\lambda_i)^{-1}\right]. \tag{C.9}$$

This is what leads to the introduction of the resolvent matrix $\mathbf{G}(z) = (z\mathbf{1}-\mathbf{M})^{-1}$, as it is now clear that

$$\rho_\varepsilon(\lambda) = \frac{1}{N\pi}\,\mathrm{Im}\left(\mathrm{Tr}\,\mathbf{G}(z)\right) := \frac{1}{\pi}\,\mathrm{Im}\left[G(z)\right], \quad G = \frac{1}{N}\,\mathrm{Tr}\,\mathbf{G}, \tag{C.10}$$

with $z = \lambda - \mathrm{i}\varepsilon \in \mathbb{C}$. This is equivalent to defining the resolvent $G(z)$ as the Stieltjes transform of the eigenvalue density, as

$$G(z) = \int \mathrm{d}\lambda\, \frac{\rho(\lambda)}{z-\lambda}. \tag{C.11}$$

The formal definition of the eigenvalue density reads thus

$$\rho(\lambda) = \frac{1}{\pi}\lim_{\varepsilon\to 0}\mathrm{Im}\left[G(\lambda-\mathrm{i}\varepsilon)\right], \tag{C.12}$$

and the knowledge of the resolvent for $z \in \mathbb{C}\setminus\mathbb{R}$ is sufficient to understand the behaviour of the eigenvalue density for large N. Note however that this method only allows to find the *continuous* part of the spectrum, and not the point-like contributions that have a vanishing contribution $\propto \frac{1}{N}$ in this framework (see below).

C.1.2 *The Wigner semi-circle distribution*

Having now introduced the resolvent G we now try to compute it for the GOE ensemble. To compute the resolvent $G(z)$ we need knowledge on the diagonal of the resolvent matrix $\mathbf{G}(z) = (z\mathbf{1} - \mathbf{M})^{-1}$. The key then is to use Schur complements, noting that with the following block matrix representation it is possible to write the inverse of a matrix as

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^t & \mathbf{C} \end{pmatrix}^{-1} = \begin{pmatrix} \left(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^t\right)^{-1} & -\left(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^t\right)^{-1}\mathbf{B}\mathbf{C}^{-1} \\ -\mathbf{C}^{-1}\mathbf{B}^t\left(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^t\right)^{-1} & \left(\mathbf{C} - \mathbf{B}^{-t}\mathbf{A}^{-1}\mathbf{B}^t\right)^{-1} \end{pmatrix}. \tag{C.13}$$

We can then apply this with the following block representation

$$z\mathbf{1} - \mathbf{M} = \begin{pmatrix} z - M_{11} & \vec{M}^t_{1,2:} \\ \vec{M}_{2:,1} & z\mathbf{1} - \mathbf{M}^{(1)} \end{pmatrix}, \tag{C.14}$$

where $\vec{M}^t_{1,2:}$ is the first line of $\mathbf{M}$ without M_{11}, $\vec{M}_{2:,1}$ the same for the first column and $\mathbf{M}^{(1)}$ the sub-matrix of $\mathbf{M}$ with the first line and column removed.

Using Schur complements, we can then write for $G_{11}(z) = (\mathbf{G}(z))$

$$G_{11}(z) = \frac{1}{z - M_{11} - \sum_{1<i,j\leq N} M_{1i} G^{(1)}_{ij}(z) M_{j1}}, \tag{C.15}$$

where $G^{(1)}_{ij} = \left(\left[z\mathbf{1} - \mathbf{M}^{(1)}\right]^{-1}\right)_{ij}$. The off-diagonal terms $G_{1j}(z)$ are given by

$$G_{1j}(z) = \frac{\sum_{1<k\leq N} M_{1k} G^{(1)}_{kj}}{z - M_{11} - \sum_{1<k,l\leq N} M_{1k} G^{(1)}_{kl}(z) M_{l1}}. \tag{C.16}$$

We then assume that the resolvent matrix is *self-averaging*, and so that the matrix $\mathbf{G}(z)$ converges to a certain matrix for large N. As there is no difference between the different indices, the diagonal terms all converge to the same quantity, i. e. $G_{11}(z) \approx G(z)$ and the off-diagonal terms all converge to a different quantity. Because of this, the off-diagonal term $G_{1j}(z)$ in Eq. (C.16) should go to 0 as $N \to \infty$, since $\mathbb{E}[M_{1k}] = 0$, and because of the argument above this should be the case of $G^{(1)}_{ij}(z)$.

It follows then that

$$\sum_{1<i,j\leq N} M_{1i} G^{(1)}_{ij}(z) M_{j1} \approx \sum_{i>1} M^2_{1i} G(z) \approx \frac{\sigma^2}{2} G(z), \tag{C.17}$$

and so Eq. (C.15) becomes

$$G(z) = \frac{1}{z - \frac{\sigma^2}{2} G(z)}, \tag{C.18}$$

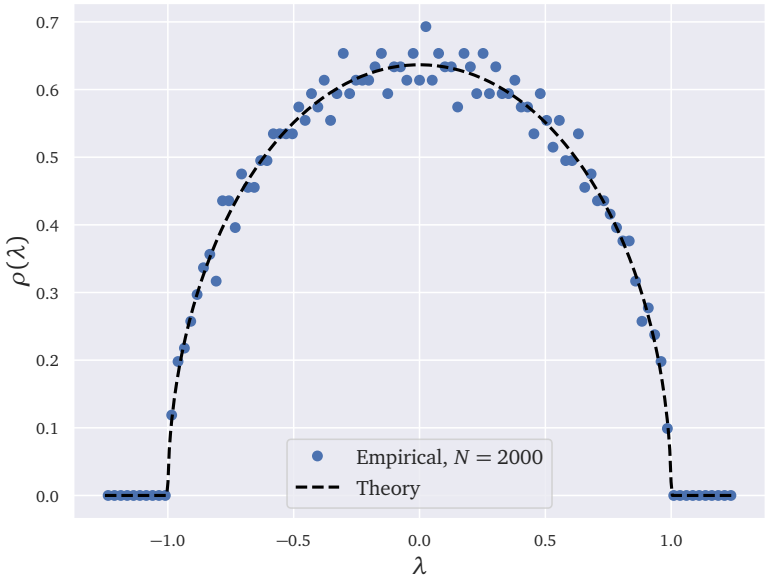


Figure C.1: A realization of a GOE matrix with $\sigma = \sqrt{2}/2$, its empirical eigenvalue distribution and the corresponding Wigner semi-circle of radius 1. Already for $N = 2000$ the agreement with theory is excellent.

as we've replaced M_{11} by its average.

This equation solves as $G(z) = \frac{z \pm \sqrt{z^2 - 2\sigma^2}}{\sigma^2}$, but it should also satisfy $G(z) \underset{|z| \gg 1}{\approx} 1/z$, and therefore

$$G(z) = \frac{z - \sqrt{z^2 - 2\sigma^2}}{\sigma^2}. \tag{C.19}$$

To find the density, we note simply that

$$\mathrm{Im}\,(G(\lambda - \mathrm{i}\varepsilon)) \approx -\frac{\varepsilon}{\sigma^2} + \frac{\sqrt{2\sigma^2 - \lambda}}{\sigma^2} \mathbf{1}\left(|\lambda| < \sigma\sqrt{2}\right), \tag{C.20}$$

giving then the celebrated Wigner semi-circle distribution

$$\rho(\lambda) = \frac{\sqrt{2\sigma^2 - \lambda}}{\pi\sigma^2} \mathbf{1}\left(|\lambda| < \sigma\sqrt{2}\right). \tag{C.21}$$

A verification of this is given in Fig. C.1, but the main message that should be retained is that the eigenvalues are spread-out over an interval centred at 0, since $\mathbb{E}[M_{ii}] = 0$, and of width $\propto \sigma$. The larger σ gets, the wider the interval.

Note that dropping the assumption of symmetric entries on N gives qualitatively the same result, except that the eigenvalues are complex. The corresponding distribution is known as Girko's circular law [132], and the eigenvalues are uniformly distributed in the complex plane within a disk of radius σ. This is shown in Fig. C.2.

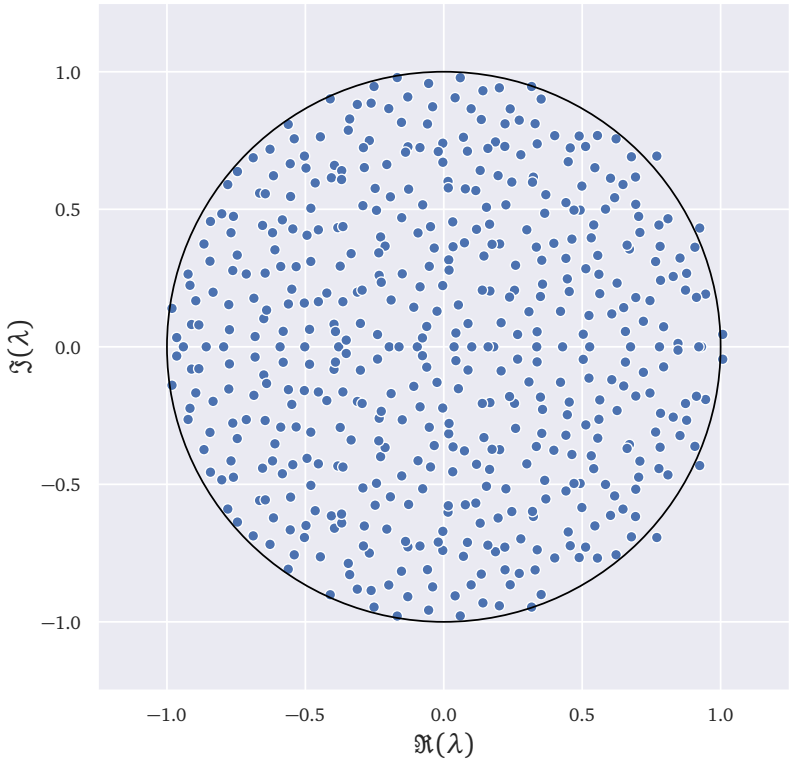


Figure C.2: Girko's circular law for a real gaussian non-symmetric matrix with entries having a variance $\sigma^2 = \frac{1}{N}$. The blue dots correspond to the spectrum of a 500×500 matrix, and the black circle has radius 1 and is centered at 0.

C.1.3 *McKay's law for the spectrum of Random Regular Graphs*

A similar case to the eigenvalue density of the adjacency matrix corresponding to an Erdös-Renyi graph, such as the one defined in Eq. (C.4), is that of the eigenvalue spectrum of a Random Regular Graph. In this graph, the nodes are labelled $i = 1, \ldots, N$, each node is connected to exactly c randomly chosen neighbours. As for the Erdös-Renyi graph, the probability of having a loop of size l scales as $\left(\frac{c}{N}\right)^l$ and goes to 0 for large N.

The matrix $\mathbf{M}$ under consideration corresponds then to the adjacency matrix of this graph, and its coefficients are either 0 or 1. First off, it is evident that

$$\mathbf{M}\vec{1} = c\vec{1}, \tag{C.22}$$

where $\vec{1} = (1, \ldots, 1)$. There is therefore an eigenvalue equal to c, with an associated eigenvector $\vec{1}$. To find the other eigenvalues, corresponding to the continuous part of the spectrum, we use again resolvent techniques.

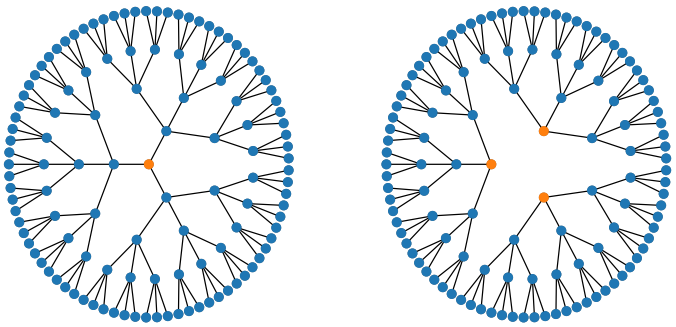


Figure C.3: Sketch of the cavity method for a Random Regular Graph with $c = 4$ neighbours. Owing to the asymptotic absence of loops, its spectrum is equivalent to that of the *Cayley tree*, also known in statistical mechanics as the *Bethe lattice*. The term $G_{11}(z)$ in the cavity equations corresponds to the central node (coloured in orange), while the term $\Delta(z)$ corresponds to the

Here, the parallel to the *cavity method* used in statistical mechanics becomes clearer.

We take again Eq. (C.15), supposing that $G_{11}(z)$ is again self-averaging, and we rewrite it as

$$G(z) = \frac{1}{z - \sum_{1<i,j\leq N} M_{1i}\Delta(z)M_{j1}} = \frac{1}{z - c\Delta(z)}, \tag{C.23}$$

where now $\Delta(z)$ corresponds to the self-averaged term coming from the Schur complement of the matrix consisting of the same graph but where one node has been removed. Note that we used that there are only c non-zero terms $M_{1i}M_{j1}$.

This situation is sketched in Fig. C.3. It is then necessary to spell out the equation verified by $\Delta(z)$, but this is clear from the structure depicted in the right hand Cayley tree on Fig. C.3: after removing the central node, connected to 4 nodes in that case, we end up with 3 normal trees. If we remove the roots of the trees, we end up with 3 identical trees for each root, and so on. The condition they verify is therefore

$$\Delta(z) = \frac{1}{z - (c-1)\Delta(z)}, \tag{C.24}$$

which we can again solve to find

$$\Delta(z) = \frac{z - \left(z^2 - 4(c-1)\right)^{1/2}}{2(c-1)}, \tag{C.25}$$

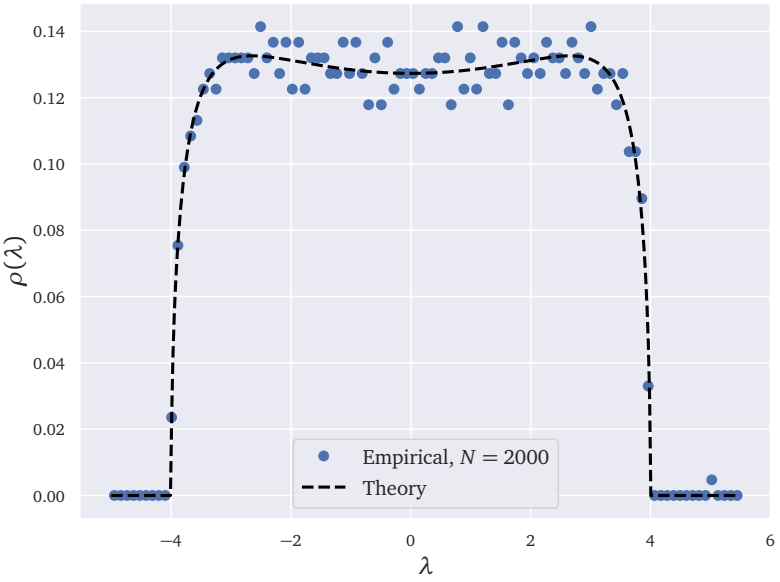


Figure C.4: Spectrum of a random regular graph with $c = 5$ neighbours per node. The dashed line shoes the theoretical value, given by Eq. (C.27), showing excellent agreement with the empirical values. Note also the single isolated eigenvalue at $\lambda = 5$ corresponding to the eigenvector $(1, \dots, 1)$.

where we've chosen the solution to Eq. (C.24) so that $\Delta(z) \underset{|z|\gg 1}{\approx} \frac{1}{z}$. Plugging into Eq. (C.23) leads to the solution

$$G(z) = \frac{z(c-2) - c\sqrt{z^2 - 4(c-1)}}{2\left(c^2 - z^2\right)}. \tag{C.26}$$

As for the Wigner semi-circle law, we can get the eigenvalue density easily by computing $G(\lambda - \mathrm{i}\varepsilon)$ for small ε. This leads now to

$$\rho(\lambda) = \frac{c\sqrt{4(c-1) - \lambda^2}}{2\pi\left(c^2 - \lambda^2\right)} \mathbf{1}\left(|\lambda| < 2\sqrt{c-1}\right), \tag{C.27}$$

also known as McKay's law, which he initially obtained using combinatorial arguments [181]. Note also that the spectrum of $\mathbf{M}/2\sqrt{c-1}$, given by $\rho\left(\frac{\lambda}{2\sqrt{c-1}}\right)$ in Eq. (C.27), converges to the semi-circle law (with $\sigma = \sqrt{2}/2$) for large c.

We show the corresponding distribution on Fig. C.4. Note that in this case the continuous part of the spectrum is again contained within an interval of length $\propto \sqrt{c-1}$, showing that more connected graphs have a more extended spectrum.

Note that it is also possible to extend this formalism to study real sparse matrices that are not necessarily symmetric, using a technique known as hermitization [227, 184], with the qualitative result remaining the same

as for Girko's law, namely that the support of the continuous part of the spectrum of $\mathbf{M}$ grows with the average connectivity c. The cavity equations (C.23,C.24) can also be solved numerically in more complicated cases using population dynamics algorithms to determine, e. g. , the spectrum of adjacency matrices supplemented by a disordered term on the diagonal, as done in Ref. [48].

C.2 THE DYSON BROWNIAN MOTION

Another way to understand this behaviour comes from using the so-called Dyson brownian motion [105]. The idea is to take an initial matrix $\mathbf{M_0}$ and to perturb it as $\mathbf{M} = \mathbf{M}_0 + \varepsilon\mathbf{V}$, with $\varepsilon \ll 1$. If we consider the eigenvalues $\left(\lambda_{i,0}\right)_{i=1\ldots N}$ and eigenvectors $\left(\left|v_{i,0}\right\rangle\right)_{i=1\ldots N}$.[1] of $\mathbf{M}_0$, then we can write for the new eigenvalues and eigenvectors λ_i' and $\left|v_i'\right\rangle$ a perturbative ansatz of the form

$$(\mathbf{M}_0 + \varepsilon\mathbf{V})\left|v_i'\right\rangle = \left(\lambda_{i,0} + \varepsilon\lambda_{i,1} + \varepsilon^2\lambda_{i,2}\ldots\right)\left(\left|v_{i,0}\right\rangle + \varepsilon\left|v_{i,1}\right\rangle + \varepsilon^2\left|v_{i,2}\right\rangle\ldots\right), \tag{C.28}$$

and after expanding in ε and grouping terms of similar degree (as is standard in perturbation theory), we end up with the following expressions for λ_i' and $\left|v_i'\right\rangle$,

$$\begin{aligned}\lambda_i' &= \lambda_{i,0} + \varepsilon\left\langle v_{i,0}\middle|\mathbf{V}\middle|v_{i,0}\right\rangle + \varepsilon^2\sum_{j(\neq i)}\frac{\left\langle v_{j,0}\middle|\mathbf{V}\middle|v_{i,0}\right\rangle^2}{\lambda_{i,0}-\lambda_{j,0}} + \mathcal{O}(\varepsilon^2)\\ \left|v_i'\right\rangle &= \left|v_{i,0}\right\rangle + \varepsilon\sum_{j(\neq i)}\frac{\left\langle v_{j,0}\middle|\mathbf{V}\middle|v_{i,0}\right\rangle}{\lambda_{i,0}-\lambda_{j,0}}\left|v_{j,0}\right\rangle + \mathcal{O}(\varepsilon).\end{aligned} \tag{C.29}$$

The two equations above have a crucial interpretation. The first line tells us that, at second order in ε, the strongest contribution to the new eigenvalue λ_i' comes mainly from the other eigenvalues λ_j that were "close" to λ_i. The second lines shows that the new eigenvector $\left|v_i'\right\rangle$ has new contribution from modes $\left|v_j\right\rangle$ with eigenvalues λ_j close to λ_i. This will be crucial later for a heuristic understanding of localization.

Having found the perturbative evolution in Eq. (3.107), we imagine the temporal dynamics of a matrix $\mathbf{M}$ as follows:

$$\mathbf{M}(t + \mathrm{d}t) = \mathbf{M}(t) + \mathrm{d}t\boldsymbol{\eta}(t), \tag{C.30}$$

where η is a random GOE matrix whose components satisfy $\mathbb{E}\left[\eta_{ij}(t)\right] = 0$ and $\mathbb{E}\left[\eta_{ij}(t)\eta_{kl}(t')\right] = \frac{\sigma^2}{2N}\left(\delta_{il}\delta_{jk} + \delta_{ik}\delta_{jl}\right)$. Note that this is equivalent

1 We use Dirac bra-ket notation due to the proximity of this subject with quantum mechanical perturbation theory.

to writing it in a "Langevin form", as $\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}t} = \boldsymbol{\eta}$, but the way it's written above is more explicit for the perturbative treatment we introduce.

Applying then Eq. (C.29) with $\varepsilon\mathbf{V} = \mathrm{d}t\,\boldsymbol{\eta}(t)$, and denoting $\lambda_i(t)$ the i-th eigenvalue of $\mathbf{M}(t)$ with its corresponding eigenvector $\left|v_i(t)\right\rangle$, yields

$$\begin{aligned}\lambda_i(t+\mathrm{d}t) &= \lambda_i(t) + \left\langle v_i(t)\middle|\boldsymbol{\eta}(t)\middle|v_i(t)\right\rangle \mathrm{d}t + \sum_{j(\neq i)} \frac{\left\langle v_j(t)\middle|\boldsymbol{\eta}(t)\middle|v_i(t)\right\rangle^2}{\lambda_i(t)-\lambda_j(t)} (\mathrm{d}t)^2 \\ &\approx \lambda_i(t) + \frac{\sigma^2}{N}\eta_i(t)\mathrm{d}t + \frac{\sigma^2}{2N}\sum_{j(\neq i)} \frac{1}{\lambda_i(t)-\lambda_j(t)}\mathrm{d}t,\end{aligned} \tag{C.31}$$

where we've used that the rotated component (since the matrix $\mathbf{O} = \sum_i |v_i\rangle\langle v_i|$ is orthogonal) $\left\langle v_i(t)\middle|\boldsymbol{\eta}\middle|v_i(t)\right\rangle$ is statistically equivalent to a random gaussian number $\frac{\sigma^2}{N}\eta_i(t)$ with

$$\mathbb{E}\left[\eta_i(t)\right] = 0, \quad \mathbb{E}\left[\eta_i(t)\eta_j(t')\right] = \delta_{ij}\delta(t-t') \tag{C.32}$$

because of the rotational invariance of the GOE ensemble. The term proportional to $(\mathrm{d}t)^2$ has also been substituted with its variance, as standard with an Itô-like change of variables (see Appendix B for more details).

Hence, we obtain the following Langevin equation for the eigenvalues λ_i,

$$\frac{\mathrm{d}\lambda_i}{\mathrm{d}t} = \frac{\sigma^2}{N}\eta_i(t) + \frac{\sigma^2}{2N}\sum_{j(\neq i)} \frac{1}{\lambda_i(t)-\lambda_j(t)}. \tag{C.33}$$

We can then consider the resolvent $G(z,t) = \frac{1}{N}\mathrm{Tr}\left[(z\mathbf{1}-\mathbf{M}(t))^{-1}\right]$, and use the Itô formula to deduce its evolution, as

$$\mathrm{d}G = \partial_z G \mathrm{d}z + \sum_i \frac{\partial G}{\partial \lambda_i}\mathrm{d}\lambda_i + \frac{1}{2}\sum_i \frac{\partial^2 G}{\partial \lambda_i^2}(\mathrm{d}\lambda_i)^2 \tag{C.34}$$

which then simply becomes

$$\begin{aligned}\frac{\mathrm{d}G(z,t)}{\mathrm{d}t} =& -\frac{\sigma^2}{2N^2}\frac{\partial}{\partial z}\left[\sum_i \frac{1}{z-\lambda_i}\sum_{j(\neq i)}\frac{1}{\lambda_i-\lambda_j}\right] + \frac{\sigma^2}{2N}\frac{\partial^2 G}{\partial z^2} \\ &+ \frac{1}{N}\sum_i \frac{1}{(z-\lambda_i)^2}\eta_i(t).\end{aligned} \tag{C.35}$$

This can be explicitly computed, and to leading order in $1/N$ gives the following Burgers equation for the resolvent,

$$\frac{\partial G(z,t)}{\partial t} + \frac{\sigma^2}{2}G(z,t)\frac{\partial G(z,t)}{\partial z} = 0. \tag{C.36}$$

This equation may then be solved using characteristics, with a solution that can be expressed with a certain function $G_0(\cdot) = G(\cdot,0)$ as

$$G(z,t) = G_0\left(z - \frac{\sigma^2 t}{2}G(z,t)\right), \tag{C.37}$$

which if we take $\mathbf{M}(0) = 0$, i. e. $G(z,0) = \frac{1}{z}$, leads to a solution [2]

$$G(z,t) = \frac{z - \sqrt{z^2 - 2\sigma^2 t}}{\sigma^2 t} = \frac{z - \sqrt{z - \sigma\sqrt{2t}}\sqrt{z + \sigma\sqrt{2t}}}{\sigma^2 t}. \quad \text{(C.38)}$$

As before, we can find the density noticing that $G(\lambda - \mathrm{i}\varepsilon, t)$ acquires a non-vanishing imaginary part in the limit $\varepsilon \to 0$ only when the argument in the square root above is negative. This leads again to a Wigner semi-circle law, but where the support of the distribution expands as $\sigma\sqrt{2t}$,

$$\rho(\lambda) = \frac{1}{\pi}\lim_{\varepsilon\to 0} G(\lambda - i\varepsilon) = \frac{1}{\pi\sigma^2 t}\sqrt{2\sigma^2 t - \lambda^2}\,\mathbf{1}\left(|\lambda| \le \sigma\sqrt{2t}\right). \quad \text{(C.39)}$$

This is, of course, a natural consequence of the fact that for long times the M_{ij} terms are gaussians with variances $\propto \frac{\sigma^2 t}{N}$. This approach also allows to find the *joint* distribution of the eigenvalues, as we show below.

C.2.1 *Joint eigenvalue distribution*

Under equation (C.30) it is clear that the variance of $M_{ij}(t)$ should scale as $\sqrt{t}$, and therefore that the eigenvalues, whose evolution is spelled out in Eq. (C.33), do not settle down into a stationary state. We can instead modify the model as

$$\mathbf{M}(t + \mathrm{d}t) = (1 - \mathrm{d}t)\,\mathbf{M}(t) + \mathrm{d}t\,\boldsymbol{\eta}(t), \quad \text{(C.40)}$$

in which case the Langevin equation for the eigenvalues becomes

$$\frac{\mathrm{d}\lambda_i}{\mathrm{d}t} = -\lambda_i + \frac{\sigma^2}{N}\eta_i(t) + \frac{\sigma^2}{2N}\sum_{j(\neq i)}\frac{1}{\lambda_i(t) - \lambda_j(t)}. \quad \text{(C.41)}$$

We can now write a Fokker-Planck equation for the density of eigenvalues $P(\{\lambda_i\}, t)$, that reads

$$\begin{aligned}\frac{\partial P(\{\lambda_i\},t)}{\partial t} =& -\frac{\sigma^2}{2N}\sum_i\left(\frac{\partial}{\partial\lambda_i}\left[\sum_{j(\neq i)}\frac{P(\{\lambda_i\},t)}{\lambda_i - \lambda_j}\right]\right) \\ &+ \sum_i\frac{\partial}{\partial\lambda_i}\left(\lambda_i P(\{\lambda_i\},t)\right) + \frac{\sigma^2}{2N}\sum_i\frac{\partial^2 P(\{\lambda_i\},t)}{\partial\lambda_i^2} \quad \text{(C.42)} \\ =& \vec{\nabla}\cdot\left[\vec{\nabla}U P(\{\lambda_i\},t)\right] + \frac{\sigma^2}{2N}\sum_i\frac{\partial^2 P(\{\lambda_i\},t)}{\partial\lambda_i^2},\end{aligned}$$

with the potential

$$U(\{\lambda_i\}) = -\frac{\sigma^2}{2N}\sum_{i<j}\log\left(|\lambda_i - \lambda_j|\right) + \sum_i\frac{\lambda_i^2}{2}. \quad \text{(C.43)}$$

2 Notice again that we pick the branch corresponding to $G(z,t) \approx 1/z$ for $|z| \gg 1$.

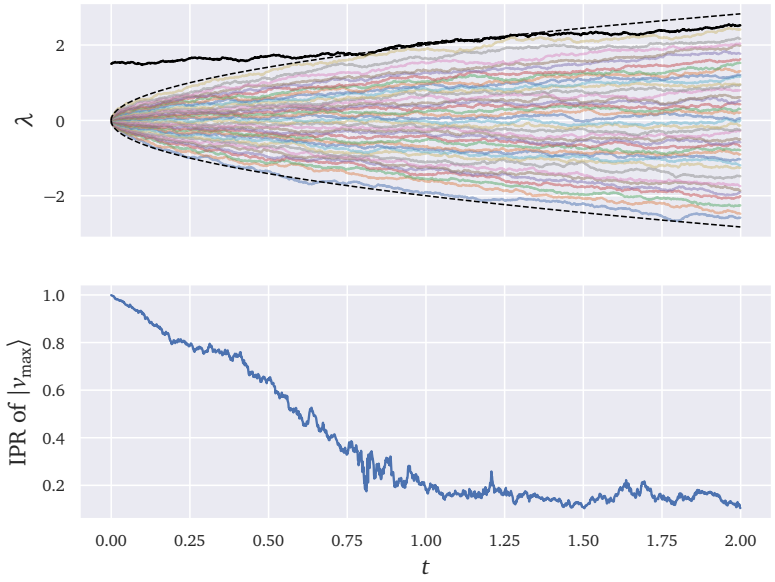


Figure C.5: Dyson brownian motion simulation inspired from [8], with $\sigma^2 = 2$. We take an initial matrix $\mathbf{M}(0) = \mathrm{diag}\,(1.5, 0, \ldots, 0)$ of size 40×40 and perturb it with GOE matrices. The last $N-1$ eigenvalues diffuse as fermions in a Coulomb potential, and therefore repulse each other. The bulk absorbs the top eigenvalue at $t^\star \approx 1$, at which point the corresponding eigenvector $|v_{\max}\rangle$ is well delocalized, having hybridized itself with modes from the bulk, as shown by its decreasing IPR value $\sum_i \left(v^i_{\max}\right)^4$, with $v^i_{\max}$ the i-th component of the vector.

In this case, the equation becomes a strict analogue of Eq. (B.14), and the stationary distribution is finally given by

$$P(\{\lambda_i\}) \propto \prod_{i<j} \left|\lambda_i - \lambda_j\right| \exp\left(-\frac{N}{\sigma^2}\sum_k \lambda_k^2\right), \tag{C.44}$$

and the λ_i eigenvalues can be interpreted as the positions of N fermions with repulsive Coulomb interactions in a harmonic trap.

D

PRODUCTION FUNCTION BASICS

D.1 CES PRODUCTION FUNCTIONS

Of standard usage in economics, the constant elasticity of substitution (CES) functions are a family of production functions giving the total production of a firm i given inputs Q_{ij} from its suppliers [1]. In the most general setting, the CES production function is defined as

$$\pi_i = z_i \left(\sum_j w_{ij} \left(\frac{Q_{ij}}{J_{ij}} \right)^{-1/q} \right)^{bq} \tag{D.1}$$

where z_i is the productivity level of firm i, the w_{ij}s are weight coefficients satisfying $\sum_j w_{ij} = 1$ and the J_{ij} terms are stoichiometric coefficients defining the number of inputs from j required to make an unit of i's output. In the remaining terms, b is called the returns to scale: multiplying all inputs Q_{ij} by some coefficient K will make the whole output level to be multiplied by K^b. In the main body and all that follows we have chosen $b = 1$, corresponding to the so-called constant returns to scale case, but our analysis can be extended to other values of b. The effect of q, the degree of substitutability, deserves a more in-depth discussion through the study of the limits $q \to \infty$ and $q \to 0$, corresponding to the so-called Cobb-Douglas and Leontief production functions.

PERFECT SUBSTITUTABILITY: COBB-DOUGLAS CASE take indeed the limit $q \to \infty$ as

$$\begin{aligned} \pi_i / z_i &= \exp\left(-q \log\left(\sum_j w_{ij} \exp\left(-\frac{1}{q} \log\left(\frac{Q_{ij}}{J_{ij}} \right) \right) \right) \right) \\ &\simeq \exp\left(-q \log\left(1 - \frac{1}{q} \sum_j w_{ij} \log\left(\frac{Q_{ij}}{J_{ij}} \right) \right) \right) \\ &\simeq \exp\left(\sum_j w_{ij} \log\left(\frac{Q_{ij}}{J_{ij}} \right) \right) = \prod_j \left(\frac{Q_{ij}}{J_{ij}} \right)^{w_{ij}} \end{aligned} \tag{D.2}$$

and one retrieves the ubiquitous Cobb-Douglas production function. In this setting, one can easily check that if a given input Q_{ij} from a firm j is multiplied by an amount $f < 1$ then the output need not drop if any other input any other input Q_{ik} is multiplied by by $f^{-\frac{w_{ik}}{w_{ij}}}$.

1 This may also include labour inputs, which we conventionally choose to correspond to the index $j = 0$.

NON-SUBSTITUTABLE INPUTS: LEONTIEF CASE take instead the limit $q \to 0$, and consider $j^* = \arg\min_j \left(\frac{Q_{ij}}{J_{ij}}\right)$ to get

$$\begin{aligned}\pi_i/z_i &= \left(w_{ij^*}\left(\frac{Q_{ij^*}}{J_{ij^*}}\right)^{-\frac{1}{q}} + \sum_{j\neq j^*} w_{ij^*}\left(\frac{Q_{ij}}{J_{ij}}\right)^{-\frac{1}{q}}\right)^{-q} \\ &= \frac{Q_{ij^{*b}}}{J_{ij^{*b}}}\left(w_{ij^*} + \sum_{j\neq j^*} w_{ij}\left(\frac{J_{ij^*}Q_{ij}}{Q_{ij^*}J_{ij}}\right)^{-\frac{1}{q}}\right)^{-q} \\ \xrightarrow[q\to 0]{} \frac{Q_{ij^*}}{J_{ij^*}} &= \min_j\left(\frac{Q_{ij}}{J_{ij}}\right)\end{aligned} \tag{D.3}$$

where the total output of firm i is determined by its scarcest input.

The CES production function therefore bridges these two limiting cases of which only the Cobb-Douglas case has been studied in the network literature. We will now study the competitive equilibrium equations for values of $q \in [0;\infty)$.

D.2 COMPETITIVE EQUILIBRIUM EQUATIONS AND HAWKINS-SIMON CONDITION

Our problem is first to maximize the profits

$$\mathscr{P}_i = \pi_i p_i - \sum_j Q_{ij} p_j \tag{D.4}$$

for each firm subject to the constraint given by Eq. (D.1). Computing first the derivative of π_i w.r.t. $Q_{il=j}$ and substituting using Eq. (D.1) yields

$$\frac{\partial \pi_i}{\partial Q_{ij}} = z_i^{-\frac{1}{q}} w_{ij} J_{ij}^{\frac{1}{q}} Q_{ij}^{-\frac{1+q}{q}} \pi_i^{\frac{1+q}{q}} \tag{D.5}$$

which can now be used to set $\frac{\partial \mathscr{P}_i}{\partial Q_{ij}} = 0$, i.e.

$$\begin{aligned} p_i z_i^{-\frac{1}{q}} w_{ij} J_{ij}^{\frac{1}{q}} Q_{ij}^{-\frac{1+q}{q}} \pi_i^{\frac{1+q}{q}} &= p_j \\ Q_{ij} &= \left(\frac{p_i}{p_j} w_{ij}\right)^{\frac{q}{1+q}} J_{ij}^{\frac{1}{q+1}} z_i^{-\frac{1}{1+q}} \pi_i. \end{aligned} \tag{D.6}$$

One can now check this solution in the Leontief and Cobb-Douglas limiting cases

$$Q_{ij}\begin{cases} \underset{q\to\infty}{=} & \frac{p_i}{p_j} w_{ij}\pi_i \\ \underset{q\to 0}{=} & J_{ij}\frac{\pi_i}{z_i} \end{cases} \tag{D.7}$$

retrieving the condition $Q_{ij} = J_{ij}\gamma_i$ defined in Eq. (5.7) of the main body, where $\gamma_i := \pi_i/z_i$ is the firm's output level.

In the Leontief case the optimal input is necessarily determined by the firm's desired output level, while it is determined by the input's price in the Cobb-Douglas case.

One needs now to impose a competitive equilibrium by setting all optimized profits to 0 as

$$\begin{aligned} \pi_i p_i &= \sum_j Q_{ij} p_j \\ \pi_i \left(z_i^{\frac{1}{1+q}} p_i^{\frac{1}{1+q}} \right) &= \pi_i \left(\sum_j w_{ij}^{\frac{q}{1+q}} J_{ij}^{\frac{1}{1+q}} p_j^{\frac{1}{1+q}} \right) \end{aligned} \tag{D.8}$$

corresponding to Eq. (5.14). The existence of a solution $p_i > 0$ for this equation is equivalent to saying that the matrix $(\widehat{\mathbf{M}})_{ij} = z_i^{\zeta} \delta_{ij} - w_{ij}^{q\zeta} J_{ij}^{\zeta}$, with $\zeta = 1/(1+q)$, is a so-called M-matrix. Once prices are determined, imposing market clearing $\pi_i = \sum_j Q_{ji}$ also leads to a similar equation

$$z_i \gamma_i - \sum_j \left(\frac{p_j}{p_i} w_{ji} \right)^{q\zeta} J_{ji}^{\zeta} z_j^{q\zeta} \gamma_j = Q_{j0} \tag{D.9}$$

which would require the matrix $\left(\widetilde{\mathbf{M}}\right)_{ij} = z_i \delta_{ij} - \left(\frac{p_j}{p_i} w_{ji}\right)^{q\zeta} J_{ji}^{\zeta} z_j^{q\zeta}$ to be an M-matrix. In the Leontief $q \to 0$ case we have the relation $\widetilde{\mathbf{M}} =^t \widehat{\mathbf{M}}$ and the sufficient condition to have both a competitive zero profit equilibrium and market clearing is that $\widehat{\mathbf{M}}$ be an M-matrix.

On the other hand, in the Cobb-Douglas case, it is easy to see that $(\widehat{\mathbf{M}})_{ij} = \delta_{ij} - w_{ij}$ is always an M-matrix, and thus the prices are chosen so that $\left(\widetilde{\mathbf{M}}\right)_{ij} = z_i \delta_{ij} - \frac{p_j}{p_i} w_{ji} z_j$ is an M-matrix to have a market-clearing equilibrium.

D.3 FUNCTIONAL ECONOMIES

In [142, 143] the authors proved that a necessary and sufficient condition for an equation such as Eq. (5.11) to have solutions is that all of the principal minors of the matrix $\mathbf{M}$ be positive, while the authors in [112] proved that this condition is equivalent to all of the eigenvalues of $\mathbf{M}$ have a positive real part.

Regarding the positivity of all principal minors of $\mathbf{M}$, Hawkins and Simon gave an economic interpretation of this condition by claiming that it is equivalent to saying that "the group of industries corresponding to each minor must be capable of supplying more than its own needs for the group of products produced by this group of industries"[143]. In this section we will provide a rewording of this in terms of an effective medium equation/Schur complements.

D.3.1 *Effective medium for one firm*

Consider first a firm i satisfying Eq. (5.10) written as

$$\mathbf{M}|P\rangle = |V\rangle \tag{D.10}$$

and consider the matrix $\mathbf{M}^{(i)}$ with row and column i removed, as well as $\left|P^{(i)}\right\rangle$ and $\left|V^{(i)}\right\rangle$ the vectors with i removed, $|J_{i\leftarrow}\rangle = (J_{i0},\ldots,J_{iN})$ and $|J_{i\rightarrow}\rangle = (J_{0i}\ldots,J_{Ni})$. The previous equation can now be written as

$$\begin{aligned} z_i p_i - \left\langle J_{i\leftarrow}|P^{(i)}\right| J_{i\leftarrow}|P^{(i)}\right\rangle &= V_i \\ \mathbf{M}^{(i)}\left|P^{(i)}\right\rangle - p_i\,|J_{i\rightarrow}\rangle &= \left|V^{(i)}\right\rangle \end{aligned} \tag{D.11}$$

multiplying now the second line by $\left(\mathbf{M}^{(i)}\right)^{-1}$, taking the product with $\langle J_{i\leftarrow}|$ and subtracting from the first line leads to

$$\begin{aligned} &\left(z_i - \left\langle J_{i\leftarrow}|(\mathbf{M}^{(i)})^{-1}|J_{i\rightarrow}\right| J_{i\leftarrow}|(\mathbf{M}^{(i)})^{-1}|J_{i\rightarrow}\right\rangle\right) p_i \\ =&V_i + \left\langle J_{i\leftarrow}|(\mathbf{M}^{(i)})^{-1}|V^{(i)}\right| J_{i\leftarrow}|(\mathbf{M}^{(i)})^{-1}|V^{(i)}\right\rangle \end{aligned} \tag{D.12}$$

which can be interpreted as i being an unique isolated firm, albeit with an effective "renormalized" productivity,

$$\tilde{z_i} := z_i - \left\langle J_{i\leftarrow}|(\mathbf{M}^{(i)})^{-1}|J_{i\rightarrow}\right| J_{i\leftarrow}|(\mathbf{M}^{(i)})^{-1}|J_{i\rightarrow}\right\rangle < z_i. \tag{D.13}$$

The Hawkins-Simons condition implies that $\mathbf{M}^{(i)}$ must also fulfil the same conditions as $\mathbf{M}$, in particular $(\mathbf{M}^{(i)})^{-1}$ has positive components. For the whole economy to be functional, the effective productivity of each firm must be positive.

D.3.2 *Sectoral interpretation*

One can also extend this analysis to different economical sectors. Consider for simplicity two sectors, so that $\mathbf{M}$ has the following block structure:

$$\mathbf{M} := \begin{pmatrix} \mathbf{M}_1 & -\mathbf{J}_{12} \\ -\mathbf{J}_{21} & \mathbf{M}_2 \end{pmatrix} \tag{D.14}$$

where the matrix $\mathbf{M}_1$ is made of firms from sector 1 and interlinkages between them, while the $\mathbf{J}$ matrices links sector 1 with sector 2. Calling $|P\rangle = (|P_1\rangle, |P_2\rangle)$ and $|V\rangle = (|V_1\rangle, |V_2\rangle)$, we can write as in Eq. (D.11) that

$$\left(\mathbf{M}_1 - \mathbf{J}_{12}\mathbf{M}_2^{-1}\mathbf{J}_{21}\right)|P_1\rangle = |V_1\rangle + \mathbf{J}_{21}\mathbf{M}_2^{-1}\,|V_2\rangle \tag{D.15}$$

and thus the Hawkins-Simon condition is fulfilled also if both $\mathbf{M}_2$ and $\mathbf{M}_1 - \mathbf{J}_{12}\mathbf{M}_2^{-1}\mathbf{J}_{21}$ are M-matrices for any choice of sector partitioning in the economy. In other words, it is not just that $\mathbf{M}_1$ must be an M-matrix by producing enough goods for the consumption of firms in sector 1, but it must also produce enough goods for all of the other sectors.

E

TECHNICAL APPENDIX TO THE ANT RECRUITMENT MODEL

E.1 DERIVATION OF THE FOKKER-PLANCK EQUATION AND STATIONARY SOLUTION

In this Section I take the continuous limit of a Master Equation, as described in Appendix B, Section B.2.

Starting from $\mathbf{P}(k,t)$, the probability that there are k ants at source A at time t, define the continuous distribution $f(x,t)$ as:

$$f(x,t) = \lim_{N\to\infty} \sum_{k=0}^{N} \delta\left(x - \frac{k}{N}\right)\mathbf{P}(k,t), \tag{E.1}$$

which amounts to replacing $\frac{k}{N}$ by x in Eqs. (7.1) and (7.2). In this case, and to leading order in $\frac{1}{N}$, the term e.g. $W(k+1\to k)\mathbf{P}(k+1,t)$ reads:

$$\left(1-\left(x+\frac{1}{N}\right)\right)\left(\varepsilon+\frac{\mu}{N}\left(x+\frac{1}{N}\right)\right)f\left(x+\frac{1}{N},t\right). \tag{E.2}$$

I proceed similarly for all terms in the right-hand side of Eq. (7.1), and Taylor-expand the left-hand side to leading order in the time variable, to obtain:

$$\begin{aligned}\partial_t f(x,t) = &\tfrac{\varepsilon}{\Delta}\big[(x+\Delta)f(x+\Delta,t)\\ &-xf(x,t)-(1-x)f(x,t)+(1-(x-\Delta))f(x-\Delta,t)\big]\\ &+\tfrac{\mu}{\Delta^2}\big[(x+\Delta)(1-(x+\Delta))f(x+\Delta,t)\\ &+(x-\Delta)(1-(x-\Delta))f(x-\Delta,t)-2x(1-x)f(x,t)\big],\end{aligned} \tag{E.3}$$

where $\Delta = \frac{1}{N}$ for simplicity.

Next, Taylor-expand the right-hand side terms, such as e.g. $(x+\Delta)f(x+\Delta,t) \approx xf(x,t)+\Delta\partial_x[xf(x,t)]+\mathcal{O}(\Delta^2)$, to order Δ for the terms with prefactor ϵ/Δ and to order Δ^2 for the terms with prefactor μ/Δ^2. Gathering everything leads to the Fokker-Planck equation:

$$\partial_t f(x,t) = -\epsilon\partial_x[(1-2x)f(x,t)]+\mu\partial_{xx}[x(1-x)f(x,t)], \tag{E.4}$$

the same as given in Eq. (7.3). This is the equivalent of the Kramers-Moyal development done in Sec. B.2.1. This equation can be written as $\partial_t f(x,t) = \partial_x J^f(x,t)$, where J^f is the probability flux, a function such that $J^f(x)\Delta$ corresponds to the probability mass flowing from $x+\Delta$ to x. To ensure the conservation of probability in $[0;1]$, I impose $J^f = 0$ at the

boundaries, meaning that no probability mass comes in or goes out during the dynamic evolution of the process.

In other words, writing $I_f(t) = \int_0^1 \mathrm{d}x\ f(x,t)$, direct integration of Eq. (E.4) leads to $\dot{I}_f(t) = J^f(1,t) - J^f(0,t) = 0$, ensuring that $I_f(t) = 1$ at all times. Keeping the next term of order Δ only slightly alters the equation:

$$\partial_t f(x,t) = -\epsilon \partial_x \left[(1-2x) f(x,t)\right] + \partial_{xx}\left[(\mu x(1-x) + \epsilon\Delta) f(x,t)\right]. \tag{E.5}$$

Recalling now, following Sec. B.1.2, that a Fokker-Planck equation of the form

$$\partial_t p(y,t) = -\partial_y \left[a(y,t)p(y,t)\right] + \partial_{yy}\left[b(y,t)p(y,t)\right] \tag{E.6}$$

corresponds to the Itô stochastic differential equation

$$\dot{y} = a(y,t) + \sqrt{b(y,t)}\eta(t) \tag{E.7}$$

where η is a brownian white noise, one readily recovers Eq. (7.4). Physically, the 0-flux boundary condition corresponds to a reflecting boundary condition: a "wall" that prevents x from getting out of $[0;1]$.

E.1.1 *Determining the stationary solution*

Looking for a stationary solution, one sets the right-hand side of Eq. (E.4) to 0, leading to

$$\frac{f_0'(x)}{f_0(x)} = (\alpha - 1)\frac{1-2x}{x(1-x)} \quad \text{with} \quad \alpha := \frac{\varepsilon}{\mu} \tag{E.8}$$

which, after direct integration, yields $f_0(x) \propto (x(1-x))^{\alpha-1}$. Integrating for $x \in [0;1]$ allows one to find the normalisation constant in terms of the Beta function, or equivalently as a ratio of Gamma functions, to get Eq. (7.5).

E.2 CHANGING INTO THE φ VARIABLE

Obtaining Eq. (7.7) and understanding the rationale behind the change of variables of Eq. (7.6) is easier by starting from Eq. (7.4).

A change of variables $x \to \varphi(x)$ leads to a new stochastic differential equation for φ, which after applying the Itô rule for differentiation given in Sec. B.1.1, reads

$$\frac{\mathrm{d}\varphi(x)}{\mathrm{d}t} = \varepsilon(1-2x)\varphi'(x) + \mu x(1-x)\varphi''(x) + \sqrt{2\mu x(1-x)}\varphi'(x)\eta(t), \tag{E.9}$$

which is still difficult to interpret because of the dependence on x of the term in front of the white noise η.

Picking however $\varphi'(x) = \frac{1}{\sqrt{x(1-x)}}$ amounts to $\varphi(x) = \arcsin(2x-1)$ and rids us of this dependence. Computing the derivatives $\varphi' = 2/\cos\varphi$ and $\varphi'' = -4\tan\varphi/\cos^2\varphi$ and replacing in Eq. (E.9):

$$\dot{\varphi} = -(2\epsilon - \mu)\tan\varphi + \sqrt{2\mu}\eta(t), \tag{E.10}$$

which because of the equivalence between stochastic differential equations and Fokker-Planck equations discussed in Appendix E.4 leads to Eq. (7.7). As before, imposing the reflecting boundary conditions $J^g(\pm\pi/2, t) = 0$ ensures conservation of probability.

Choosing instead to keep the term of order Δ given in Eq. (E.5) leads to a Langevin equation

$$\dot{x} = \varepsilon(1-2x) + \sqrt{2\mu x(1-x) + 2\varepsilon\Delta}\eta(t), \tag{E.11}$$

which motivates the change of variables

$$\phi = \arctan\left(\frac{2x-1}{2\sqrt{x(1-x)+\alpha\Delta}}\right), \tag{E.12}$$

where now $|\phi| \leq \arctan\left(1/\sqrt{2\alpha\Delta}\right) \approx \frac{\pi}{2} - 2\sqrt{\alpha\Delta}$, and naturally one can check that the definition of ϕ corresponds to φ as $\Delta \to 0$, with $\phi \approx \varphi - 2\alpha\Delta\tan(\varphi)$ to leading order in Δ. The analysis in the limit $N \to \infty$ therefore holds only in the limit $\tan(\varphi) \ll \frac{N}{2\alpha}$.

This new variable actually verifies the very same SDE, Eq. (E.9), but with a different boundary.

E.3 PROPERTIES OF THE SOLUTION

I take the solutions to Eq. (7.14) as those given in [261]. I will first check that they satisfy the boundary condition.

E.3.1 *Checking the boundary condition*

I begin by reminding the reader of the following definition:

$${}_2F_1(-k, a; b, u) = \sum_{l=0}^{k}\binom{k}{\ell}(-1)^\ell \frac{\Gamma(a+\ell)}{\Gamma(a)}\frac{\Gamma(b)}{\Gamma(b+\ell)}u^\ell. \tag{E.13}$$

In this case, direct differentiation in Eq. (7.14) for e. g. even modes $n = 2k$ in the limit $\varphi \to \pm\frac{\pi}{2}$leads to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varphi}\left({}_2F_1\left(-k, \beta+k; \beta+\frac{1}{2}, \cos^2\varphi\right)\right) &= 2\sin\varphi\cos\varphi\frac{k(\beta+k)}{\beta+1/2} + \mathcal{O}(\cos\varphi) \\ &\approx \pm 2\left(\frac{\pi}{2} \mp \varphi\right)\frac{k(\beta+k)}{\beta+1/2}.\end{aligned}$$

(E.14)

With this, one can directly compute, with ${}_2F_1\left(-k,\beta+k;\beta+\frac{1}{2},1\right):=c_1$ and for $\varphi\to\pm\frac{\pi}{2}$:

$$\begin{aligned}\beta\tan\varphi\Psi_{2k}(\varphi)+\Psi'_{2k}(\varphi)\approx&\beta\tan\varphi\cos^\beta\varphi\ c_1-\beta\tan\varphi\cos^\beta\varphi\ c_1\\&\pm2\cos^\beta\varphi\ \left(\frac{\pi}{2}\mp\varphi\right)\frac{k(\beta+k)}{\beta+1/2}\\\approx&\pm2\frac{k(\beta+k)}{\beta+1/2}\left(\frac{\pi}{2}\mp\varphi\right)^{1+\beta},\end{aligned}\tag{E.15}$$

which after multiplication with $\cos^\beta\varphi\approx\left(\frac{\pi}{2}\mp\varphi\right)^\beta$ proves Eq. (7.15) for $n=2k$.

The proof for odd $n=2k+1$ is strictly equivalent. It therefore follows that the solutions of [261], although found initially for vanishing boundary conditions, also satisfy the boundary condition given in Eq. (7.12).

E.3.2 *Explicit expressions*

In this section I discuss the explicit expressions of the functions f_n and the constants A_n.

The constants A_n are set so that $\int_{-\pi/2}^{\pi/2}\Psi_n\Psi_m=\delta_{n,m}$, and therefore implies, in terms of the variable α,

$$\begin{aligned}A_{2k}(\alpha)=&\left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\mathrm{d}\varphi\ \cos^{2\alpha-1}\varphi\ {}_2F_1\left(-k,\alpha+k-\frac{1}{2};\alpha,\cos^2\varphi\right)^2\right)^{-1/2}\\A_{2k+1}(\alpha)=&\left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\mathrm{d}\varphi\ \cos^{2\alpha-1}\varphi\ \sin^2\varphi\ {}_2F_1\left(-k,\alpha+k+\frac{1}{2};\alpha,\cos^2\varphi\right)^2\right)^{-1/2}.\end{aligned}\tag{E.16}$$

To substitute and find the expressions of $f_n(x)$, we recall that

$$\sin\varphi=2x-1,\quad\cos\varphi=2\sqrt{x(1-x)}\tag{E.17}$$

and get, using Eq. (7.14) and replacing into $f_n(x)=\frac{\sqrt{g_0(\varphi(x))}\Psi_n(\varphi(x))}{2\sqrt{x(1-x)}}$, the explicit expressions

$$f_{2k}(x)=A_{2k}(\alpha)\sqrt{\frac{\Gamma(\alpha+1/2)}{\sqrt{\pi}\Gamma(\alpha)}}(4x(1-x))^{\alpha-1}\tag{E.18}$$

$${}_2F_1\left(-k,\alpha+k-\frac{1}{2};\alpha,4x(1-x)\right)\tag{E.19}$$

$$f_{2k+1}(x)=A_{2k+1}(\alpha)\sqrt{\frac{\Gamma(\alpha+1/2)}{\sqrt{\pi}\Gamma(\alpha)}}(4x(1-x))^{\alpha-1}(2x-1)\tag{E.20}$$

$${}_2F_1\left(-k,\alpha+k+\frac{1}{2};\alpha,4x(1-x)\right).\tag{E.21}$$

E.3.3 *Computing the moments of the distribution*

To understand the dynamics of the moments of the distribution

$$\mathbb{E}\left[x^m(t)\right] = \int_0^1 \mathrm{d}x\ f(x,t)x^m \tag{E.22}$$

it is necessary to understand the behaviour of $B_{n,m} = \int_0^1 \mathrm{d}x\ f_n(x)x^m$. Owing to the parity of $f_n(x)$ with respect to $x = 1/2$ it is clear that for even moments $m = 2p$ only even modes $n = 2k$ will be non zero, and vice versa for odd moments and modes.

We therefore develop the computation of even moments only, as the extension to odd moments is direct. We wish to evaluate the integral $\int_0^1 \mathrm{d}x\ f_{2k}(x)x^{2p} = 2\int_0^{1/2} \mathrm{d}x\ f_{2k}(x)x^{2p}$, after changing variables as $t = 4x(1-x)$, it is clear that this integral is proportional to

$$I_{2p,2m} = \int_0^1 \mathrm{d}t\ {}_2F_1\left(-k, \alpha+k-\frac{1}{2}; \alpha, t\right) t^{\alpha-1}(1-t)^{p-1/2}. \tag{E.23}$$

After expanding the hypergeometric function and integrating explicitly, we find

$$I_{2k,2p} = \frac{\Gamma(\alpha)\Gamma(1/2+p)}{\Gamma(\beta+k)} \sum_{l=0}^{k} \binom{k}{l} (-1)^l \frac{\Gamma(\beta+k+l)}{\Gamma(\beta+1+l+p)} \tag{E.24}$$

requiring then the explicit computation of the right-hand side sum, which we denote by $S_{2k,2p}$.

Mind that for $m = 2p+1$ the only modes that contribute are $n = 2k+1$, and the equivalent of the previous integral is

$$I_{2k+1,2p+1} = \frac{\Gamma(\alpha)\Gamma(3/2+p)}{\Gamma(\beta+k+1)} \sum_{l=0}^{k} \binom{k}{l} (-1)^l \frac{\Gamma(\beta+1+k+l)}{\Gamma(\beta+2+l+p)}, \tag{E.25}$$

allowing to define the sum $S_{2k+1,2p+1}$ as the term on the right-hand side.

We discuss this for $k \geq 1$ in two situations, $k > m$ and $k \leq m$.

E.3.3.1 *First case: $k > m$*

We can then write the sum $S_{2k,2p}$ as

$$\sum_{l=0}^{k} \binom{k}{l} (-1)^l \prod_{i=p+1}^{k-1} (\beta+l+i), \tag{E.26}$$

which, written as such, leads us to introduce the function

$$P(X) = \sum_{l=0}^{k} \binom{k}{l} (-1)^l X^{\beta+l+k-1} = X^{\beta+k-1}(1-X)^k. \tag{E.27}$$

Applying the generalized Leibniz rule to compute the $k-p-1$-th derivative of this function, we obtain directly that $S_{2k,2p} = P^{(k-p-1)}(1) = 0$ in this case. A similar calculation can be done for $S_{2k+1,2p+1}$, and it follows therefore that

$$\int_0^1 \mathrm{d}x\ f_n(x)x^m = 0 \quad \text{for } n > m. \tag{E.28}$$

E.3.3.2 *Second case:* $k \leq m$

In this case, we now write the sum as

$$\sum_{l=0}^{k}\binom{k}{l}(-1)^l\prod_{i=k}^{p}\frac{1}{\beta+l+i}, \tag{E.29}$$

which can instead be seen as the result of successive integrations on the function defined in Eq. (E.27).

To compute it, we define the functions $\mathrm{B}_0(t;a,b) = t^{a-1}(1-t)^{b-1}$ and $\mathrm{B}_{n+1}(t;a,b) = \int_0^t \mathrm{d}u\ \mathrm{B}_n(u;a,b)$, with B_1 corresponding to the standard incomplete Beta function. With this definition, the sum reads

$$S_{2k,2p} = \int_0^1 \mathrm{d}u\ \mathrm{B}_{p-k}(u;\beta+k+1,k+1), \tag{E.30}$$

while on the other hand successive integration by parts gives

$$\begin{aligned}\mathrm{B}_n(1;a,b) = &\left[\sum_{j=0}^{n-1}(-1)^{j+1}\frac{(t-1)^{j+1}}{\Gamma(j+2)}\mathrm{B}_{n-j}(u;a,b)\right]_0^1\\ &+(-1)^n\int_0^1 \mathrm{d}u\ \frac{(t-1)^n}{\Gamma(n+1)}\mathrm{B}_0(u;a,b)\\ =&\frac{\mathrm{B}_1(n+a,b)}{\Gamma(n+1)}.\end{aligned} \tag{E.31}$$

Finally, gathering everything we get

$$S_{2k,2p} = \frac{\Gamma(p+\beta+1)\Gamma(k+1)}{\Gamma(p+\beta+k+2)\Gamma(p-k+1)}, \tag{E.32}$$

while replacing $\beta \to \beta+1$ gives the similar expression

$$S_{2k+1,2p+1} = \frac{\Gamma(p+\beta+2)\Gamma(k+1)}{\Gamma(p+\beta+k+3)\Gamma(p-k+1)}. \tag{E.33}$$

The final result follows,

$$B_{n,m} = \int_0^1 \mathrm{d}x\ f_n(x)x^m = A_n(\alpha)\sqrt{\frac{\Gamma(\alpha+1/2)}{\sqrt{\pi}\Gamma(\alpha)}}I_{n,m}\mathbf{1}\,(n \leq m) \tag{E.34}$$

with

$$\begin{aligned} I_{2k,2p+1} &= 0 \\ I_{2k,2p} &= \frac{\Gamma(\alpha)\Gamma(1/2+p)\Gamma(p+\beta+1)\Gamma(k+1)}{\Gamma(\beta+k)\Gamma(p+\beta+k+2)\Gamma(p-k+1)} \\ I_{2k+1,2p+1} &= \frac{\Gamma(\alpha)\Gamma(3/2+p)\Gamma(p+\beta+2)\Gamma(k+1)}{\Gamma(\beta+k+1)\Gamma(p+\beta+k+3)\Gamma(p-k+1)}, \end{aligned} \tag{E.35}$$

allowing then for explicit computation of the dynamics of $\mathbb{E}[x^m(t)]$.

E.4 STOCHASTIC CALCULUS TECHNIQUES

In this Appendix, we shall directly integrate stochastic differential equations describing the model to obtain information on the covariances of moments $x^n(t)$. We begin by looking at the covariance $\mathrm{Cov}(x(t+T),x(T))$.

A direct integration of Eq. (7.4) leads to

$$\begin{aligned} x(t+T) =& x(T)+\varepsilon t-2\varepsilon\int_T^{t+T} \mathrm{d}s\; x(s) \\ &+\int_T^{t+T} \mathrm{d}s\; \sqrt{2\mu x(s)(1-x(s))}\eta(s). \end{aligned} \tag{E.36}$$

Taking now the covariance with $x(t)$ and using linearity,

$$\begin{aligned} \mathrm{Cov}(x(t+T),x(T)) =& \mathrm{Cov}(x(T),x(T))-2\varepsilon\int_T^{t+T} \mathrm{d}s\; \mathrm{Cov}(x(s),x(T)) \\ &+\int_T^{t+T} \mathrm{d}s\; \mathbb{E}\left[\sqrt{2\mu x(s)(1-x(s))}x(T)\eta(s)\right] \end{aligned} \tag{E.37}$$

with the last integral being equal to 0 through the independence of η, as

$$\mathbb{E}\left[\sqrt{2\mu x(s)(1-x(s))}x(T)\right]\mathbb{E}[\eta(s)]=0. \tag{E.38}$$

Taking finally the derivative with respect to t and solving the resulting differential equation we find

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}s}\mathrm{Cov}(x(T+s),x(T)) = -2\varepsilon\mathrm{Cov}(x(T+s),x(T)) \\ &\mathrm{Cov}(x(t+T),x(T)) \propto e^{-2\varepsilon t}. \end{aligned} \tag{E.39}$$

Similarly, one can derive the stochastic differential equation followed by $\sigma_2(x)=x(1-x)$ using the differentiation rule exemplified in Eq. (E.9), namely

$$\frac{\mathrm{d}[x(1-x)]}{\mathrm{d}t} = \varepsilon-(4\varepsilon+2\mu)x(1-x)+\sqrt{2\mu x(1-x)}(1-2x)\eta(t),$$

(E.40)

and as before, we can take the covariance $\mathrm{Cov}(\sigma_2(x(t+T)),\sigma_2(x(T)))$, differentiate with respect to t and find that it satisfies a differential equation, which after integrating reads

$$\mathrm{Cov}(\sigma_2(x(t+T)),\sigma_2(x(T))) \propto e^{-(4\varepsilon+2\mu)t}. \tag{E.41}$$

This method can be extended to computing correlators defined as

$$C_{n,k}(t+T,T) := \mathrm{Cov}\left(x(t+T)^n, x(T)^k\right). \tag{E.42}$$

Applying Itô calculus as before, one can show that these functions satisfy the following ODE system:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[C_{n,k}(T+s,T)\right] = -\mu\mathscr{E}_n C_{n,k}(T+s,T) + \mu n(n-1+\alpha)C_{n-2,k}(T+s,T). \tag{E.43}$$

Owing to its triangular structure, it can be diagonalized iteratively to find functions σ_n, such that $\sigma_n(x)$ is a polynomial of degree n and that the covariances $C_{\sigma_n}(T+s,T) = \mathrm{Cov}\left[\sigma_n(x(T+s)),\sigma_n(x(T))\right]$ satisfy

$$\frac{\mathrm{d}}{\mathrm{d}s}C_{\sigma_n}(T+s,T) = -\mu\mathscr{E}_n C_{\sigma_n}(T+s,T). \tag{E.44}$$

Knowing that $\sigma_1(x) = x$ and $\sigma_2(x) = x(1-x)$, it is possible to find the third combination $\sigma_3(x) = (2x-1)\left[\left(1+\frac{2\alpha}{3}\right)(2x-1)^2-1\right]$. Integrating the equations in Eq. (E.44), one finds then that

$$C_{\sigma_n}(t+T,T) \propto e^{-\mu\mathscr{E}_n t}. \tag{E.45}$$

These results can also be obtained directly from the eigenvalues and eigenfunctions of the Schrödinger problem.

BIBLIOGRAPHY

[1] Andrew Abel. *Asset Prices under Habit Formation and Catching up with the Joneses*. Tech. rep. Mar. 1990. DOI: 10.3386/w3279. URL: https://doi.org/10.3386/w3279.

[2] Daron Acemoglu, Ufuk Akcigit, and William R. Kerr. "Networks and the Macroeconomy: An Empirical Exploration". In: *SSRN Electronic Journal* (2015). DOI: 10.2139/ssrn.2630102. URL: https://doi.org/10.2139%2Fssrn.2630102.

[3] Daron Acemoglu and Asuman Ozdaglar. "Opinion Dynamics and Learning in Social Networks". In: *Dynamic Games and Applications* 1.1 (2011), pp. 3–49. ISSN: 2153-0793. DOI: 10.1007/s13235-010-0004-1.

[4] Daron Acemoglu, Vasco Carvalho, Asu Ozdaglar, and Alireza Tahbaz-Salehi. "The Network Origins of Aggregate Fluctuations". In: *Econometrica* 80.5 (2012), pp. 1977–2016. DOI: 10.3982/ecta9623. URL: https://doi.org/10.3982%2Fecta9623.

[5] Reka Albert and Albert-Laszlo Barabasi. "Statistical mechanics of complex networks". In: *Reviews of Modern Physics* 74.1 (2002). arXiv: cond-mat/0106096, pp. 47–97. ISSN: 0034-6861, 1539-0756. DOI: 10.1103/RevModPhys.74.47.

[6] Peter M Allen and Jacqueline M McGlade. "Dynamics of discovery and exploitation: the case of the Scotian Shelf groundfish fisheries". In: *Canadian Journal of Fisheries and Aquatic Sciences* 43.6 (1986), pp. 1187–1200.

[7] Romain Allez and Jean-Philippe Bouchaud. "Eigenvector dynamics: General theory and some applications". In: *Physical Review E* 86.4 (2012). DOI: 10.1103/physreve.86.046202. URL: https://doi.org/10.1103%2Fphysreve.86.046202.

[8] Romain Allez, Joël Bun, and Jean-Philippe Bouchaud. "The eigenvectors of Gaussian matrices with an external source". In: *arXiv:1412.7108 [cond-mat]* (2015). arXiv: 1412.7108. URL: http://arxiv.org/abs/1412.7108.

[9] Henri Alloul. *Introduction to the Physics of Electrons in Solids*. Springer Science & Business Media, 2010.

[10] Johannes Alt, Raphael Ducatez, and Antti Knowles. "Delocalization transition for critical Erdös-Renyi graphs". In: *arXiv:2005.14180 [math-ph]* (2020). arXiv: 2005.14180. URL: http://arxiv.org/abs/2005.14180.

[11] Eduardo G. Altmann and Martin Gerlach. “Statistical laws in linguistics”. In: *arXiv:1502.03296 [physics]* (2016). arXiv: 1502.03296, pp. 7–26. DOI: 10.1007/978-3-319-24403-7_2.

[12] L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, H. E. Stanley, and M. H. R. Stanley. “Scaling behavior in economics: I. Empirical results for company growth”. In: *arXiv:cond-mat/9702082* (1997). arXiv: cond-mat/9702082. DOI: 10.1051/jp1:1997180. URL: http://arxiv.org/abs/cond-mat/9702082.

[13] Ariel Amir, Naomichi Hatano, and David R. Nelson. “Non-Hermitian localization in biological networks”. In: *Physical Review E* 93.4 (2016). DOI: 10.1103/physreve.93.042310. URL: https://doi.org/10.1103%2Fphysreve.93.042310.

[14] Ariel Amir, Yuval Oreg, and Yoseph Imry. “On relaxations and aging of various glasses”. In: *Proceedings of the National Academy of Sciences* 109.6 (2012), pp. 1850–1855. ISSN: 0027-8424. DOI: 10.1073/pnas.1120147109. eprint: https://www.pnas.org/content/109/6/1850.full.pdf. URL: https://www.pnas.org/content/109/6/1850.

[15] Kartik Anand, Jonathan Khedair, and Reimer Kühn. “Structural model for fluctuations in financial markets”. In: *Physical Review E* 97.5 (2018), p. 052312. DOI: 10.1103/PhysRevE.97.052312. URL: https://link.aps.org/doi/10.1103/PhysRevE.97.052312.

[16] P. W. Anderson. “Absence of Diffusion in Certain Random Lattices”. In: *Physical Review* 109.5 (1958), pp. 1492–1505. DOI: 10.1103/physrev.109.1492. URL: https://doi.org/10.1103%2Fphysrev.109.1492.

[17] P. W. Anderson. “More Is Different”. In: *Science* 177.4047 (1972), pp. 393–396. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.177.4047.393. URL: https://science.sciencemag.org/content/177/4047/393.

[18] Simon P. Anderson, Andre de Palma, and Jacques-Francois Thisse. *Discrete Choice Theory of Product Differentiation (The MIT Press)*. The MIT Press, 1992. ISBN: 026201128X. URL: https://www.amazon.com/Discrete-Choice-Theory-Product-Differentiation/dp/026201128X?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=026201128X.

[19] Tibor Antal, Arne Traulsen, Hisashi Ohtsuki, Corina E. Tarnita, and Martin A. Nowak. “Mutation-selection equilibrium in games with multiple strategies”. In: *Journal of Theoretical Biology* 258.4 (2009), pp. 614–622. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2009.02.010.

[20] Hideaki Aoyama, Yoshi Fujiwara, Yuichi Ikeda, Hiroshi Iyetomi, and Wataru Souma. "Econophysics and companies". In: *Cambridge Books* (2010). DOI: 10.1017/CBO9780511761157.

[21] W. Brian Arthur. "Positive Feedbacks in the Economy". In: *Scientific American* 262.2 (1990), pp. 92–99. ISSN: 00368733, 19467087. URL: http://www.jstor.org/stable/24996687.

[22] W. Brian Arthur. "Inductive Reasoning and Bounded Rationality". In: *The American Economic Review* 84.2 (1994), pp. 406–411. ISSN: 00028282. URL: http://www.jstor.org/stable/2117868.

[23] W. Brian Arthur, Yu.M. Ermoliev, and Yu.M. Kaniovski. "Path-dependent processes and the emergence of macro-structure". In: *European Journal of Operational Research* 30.3 (June 1987), pp. 294–303. DOI: 10.1016/0377-2217(87)90074-9. URL: https://doi.org/10.1016/0377-2217(87)90074-9.

[24] Enghin Atalay, Ali Hortaçsu, James Roberts, and Chad Syverson. "Network structure of production". In: *Proceedings of the National Academy of Sciences* (2011). DOI: 10.1073/pnas.1015564108. eprint: https://www.pnas.org/content/early/2011/03/07/1015564108.full.pdf. URL: https://www.pnas.org/content/early/2011/03/07/1015564108.

[25] R. L. Axtell. "Zipf Distribution of U.S. Firm Sizes". In: *Science* 293.5536 (2001), pp. 1818–1820. DOI: 10.1126/science.1062081. URL: https://doi.org/10.1126%2Fscience.1062081.

[26] Robert Axtell. *Endogenous Dynamics of Multi-Agent Firms*. ID 2827059. 2015. URL: https://papers.ssrn.com/abstract=2827059.

[27] Nicolas Bacaër. "Lotka, Volterra and the predator–prey system (1920–1926)". In: *A Short History of Mathematical Population Dynamics*. Ed. by Nicolas Bacaër. Springer, 2011, pp. 71–76. ISBN: 978-0-85729-115-8. DOI: 10.1007/978-0-85729-115-8_13. URL: https://doi.org/10.1007/978-0-85729-115-8_13.

[28] L. Bachelier. "Théorie de la spéculation". In: *Annales scientifiques de l'École normale supérieure* 17 (1900), pp. 21–86. ISSN: 0012-9593, 1873-2151. DOI: 10.24033/asens.476. URL: http://www.numdam.org/item?id=ASENS_1900_3_17__21_0.

[29] Kent Bækgaard Lauritsen, Stefano Zapperi, and H. Eugene Stanley. "Self-organized branching processes: Avalanche models with dissipation". In: *Physical Review E* 54.3 (1996), pp. 2483–2488. DOI: 10.1103/PhysRevE.54.2483.

[30] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices". In: *The Annals of Probability* 33.5 (2005), pp. 1643–1697. DOI: 10.1214/009117905000000233. URL: https://doi.org/10.1214%2F009117905000000233.

[31] Per Bak. *How Nature Works*. Springer New York, 1996. DOI: 10.1007/978-1-4757-5426-1. URL: https://doi.org/10.1007/978-1-4757-5426-1.

[32] Per Bak. *How nature works: the science of self-organized criticality*. Springer Science & Business Media, 2013. DOI: 10.1007/978-1-4757-5426-1.

[33] Per Bak, Chao Tang, and Kurt Wiesenfeld. "Self-organized criticality: An explanation of the 1/ f noise". In: *Physical Review Letters* 59.4 (1987), pp. 381–384. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.59.381.

[34] Per Bak, Kan Chen, José Scheinkman, and Michael Woodford. "Aggregate Fluctuations from Independent Sectoral Shocks: Self-Organized Criticality in a Model of Production and Inventory Dynamics". In: (1992). DOI: 10.3386/w4241. URL: https://doi.org/10.3386%2Fw4241.

[35] David Baqaee and Emmanuel Farhi. *Supply and Demand in Disaggregated Keynesian Economies with an Application to the Covid-19 crisis*. 2020. URL: https://scholar.harvard.edu/farhi/publications/keynesian-production-networks-application-covid-19-crisis.

[36] Marco Bardoscia, Giacomo Livan, and Matteo Marsili. "Statistical mechanics of complex economies". In: *Journal of Statistical Mechanics: Theory and Experiment* 2017.4 (2017), p. 043401. DOI: 10.1088/1742-5468/aa6688. URL: https://doi.org/10.1088%2F1742-5468%2Faa6688.

[37] Alain Barrat and Marc Mézard. "Phase space diffusion and low temperature aging". In: *Journal de Physique I* 5.8 (1995), pp. 941–947. DOI: 10.1051/jp1:1995174.

[38] Gary S. Becker. "A Note on Restaurant Pricing and Other Examples of Social Influences on Price". In: *Journal of Political Economy* 99.5 (1991), pp. 1109–1116. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/2937660.

[39] R. Beckers, J. L. Deneubourg, S. Goss, and J. M. Pasteels. "Collective decision making through food recruitment". In: *Insectes Sociaux* 37.3 (Sept. 1990), pp. 258–267. DOI: 10.1007/bf02224053. URL: https://doi.org/10.1007/bf02224053.

[40] Eric Beinhocker et al. *Inclusive Economics Is Complexity Economics*. 2019. URL: http://bostonreview.net/forum/economics-after-neoliberalism/complexity-economists-inclusive-economics-complexity-economics.

[41] R. J. Bell and P. Dean. "Atomic vibrations in vitreous silica". In: *Discussions of the Faraday Society* 50 (1970), p. 55. DOI: 10.1039/df9705000055. URL: https://doi.org/10.1039/df9705000055.

[42] Gérard Ben Arous, Leonid V. Bogachev, and Stanislav A. Molchanov. "Limit theorems for sums of random exponentials". In: *Probability Theory and Related Fields* 132.4 (2005), pp. 579–612. ISSN: 0178-8051, 1432-2064. DOI: 10.1007/s00440-004-0406-3.

[43] Florent Benaych-Georges and Sandrine Péché. "Largest eigenvalues and eigenvectors of band or sparse random matrices". In: *Electronic Communications in Probability* 19.0 (2014). DOI: 10.1214/ecp.v19-3027. URL: https://doi.org/10.1214%2Fecp.v19-3027.

[44] Ben Bernanke, Mark Gertler, and Simon Gilchrist. *The Financial Accelerator and the Flight to Quality*. Tech. rep. 1994. DOI: 10.3386/w4789. URL: https://doi.org/10.3386%2Fw4789.

[45] E Bertin and J-P Bouchaud. "Dynamical ultrametricity in the critical trap model". In: *Journal of Physics A: Mathematical and General* 35.13 (2002), pp. 3039–3051. DOI: 10.1088/0305-4470/35/13/302. URL: https://doi.org/10.1088%2F0305-4470%2F35%2F13%2F302.

[46] G Biroli and R Monasson. "A single defect approximation for localized states on random lattices". In: *Journal of Physics A: Mathematical and General* 32.24 (1999), pp. L255–L261. DOI: 10.1088/0305-4470/32/24/101. URL: https://doi.org/10.1088%2F0305-4470%2F32%2F24%2F101.

[47] Giulio Biroli, Guy Bunin, and Chiara Cammarota. "Marginally stable equilibria in critical ecosystems". In: *New Journal of Physics* 20.8 (2018), p. 083051. DOI: 10.1088/1367-2630/aada58. URL: https://doi.org/10.1088%2F1367-2630%2Faada58.

[48] Giulio Biroli, Guilhem Semerjian, and Marco Tarzia. "Anderson Model on Bethe Lattices: Density of States, Localization Properties and Isolated Eigenvalue". In: *Progress of Theoretical Physics Supplement* 184 (2010), pp. 187–199. DOI: 10.1143/ptps.184.187. URL: https://doi.org/10.1143%2Fptps.184.187.

[49] Olivier Blanchard. "Where danger lurks". In: *Finance & Development* 51.3 (2014), pp. 28–31.

[50] V. V. Bochkarev and E. Yu Lerner. "The Zipf law for random texts with unequal probabilities of occurrence of letters and the Pascal pyramid". In: *Russian Mathematics* 56.12 (2012). arXiv: 1205.0796, pp. 25–27. ISSN: 1066-369X, 1934-810X. DOI: 10.3103/S1066369X12120031.

[51] Thierry Bochud and Damien Challet. "Optimal approximations of power-laws with exponentials". In: (2006). arXiv: physics/0605149 [physics.data-an].

[52] Julius Bonart, Jean-Philippe Bouchaud, Augustin Landier, and David Thesmar. "Instabilities in large economies: aggregate volatility without idiosyncratic shocks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2014.10 (2014), P10040. DOI: 10.1088/

1742-5468/2014/10/p10040. URL: https://doi.org/10.1088%2F1742-5468%2F2014%2F10%2Fp10040.

[53] Christian Borghesi and Jean-Philippe Bouchaud. "Of songs and men: a model for multiple choice with herding". In: *Quality & Quantity* 41.4 (Mar. 2007), pp. 557–568. DOI: 10.1007/s11135-007-9074-6. URL: https://doi.org/10.1007/s11135-007-9074-6.

[54] Massimo Borlandi. "Expliquer les régularités sociales : Durkheim critique et continuateur de Quetelet". In: *Durkheim*. Presses Universitaires de France, 2008, p. 79. DOI: 10.3917/puf.valad.2008.01.0079. URL: https://doi.org/10.3917/puf.valad.2008.01.0079.

[55] G. Bottazzi and A. Secchi. "A new class of asymmetric exponential power densities with applications to economics and finance". In: *Industrial and Corporate Change* 20.4 (2011), pp. 991–1030. ISSN: 0960-6491, 1464-3650. DOI: 10.1093/icc/dtr036.

[56] Giulio Bottazzi, Le Li, and Angelo Secchi. "Aggregate fluctuations and the distribution of firm growth rates". In: *Industrial and Corporate Change* 28.3 (2019), pp. 635–656. ISSN: 0960-6491. DOI: 10.1093/icc/dtz016.

[57] Giulio Bottazzi and Angelo Secchi. "Why are distributions of firm growth rates tent-shaped?" In: *Economics Letters* 80.3 (2003), pp. 415–420. ISSN: 0165-1765. DOI: 10.1016/S0165-1765(03)00142-3.

[58] Giulio Bottazzi and Angelo Secchi. "Explaining the Distribution of Firm Growth Rates". In: *The RAND Journal of Economics* 37.2 (2006), pp. 235–256. ISSN: 07416261. URL: http://www.jstor.org/stable/25046240.

[59] J. P. Bouchaud. "Weak ergodicity breaking and aging in disordered systems". In: *Journal de Physique I* 2.9 (Sept. 1992), pp. 1705–1713. DOI: 10.1051/jp1:1992238. URL: https://doi.org/10.1051/jp1:1992238.

[60] J. P. Bouchaud, S. Ciliberti, Y. Lempérière, A. Majewski, P. Seager, and K. Sin Ronia. *Black was right: Price is within a factor 2 of Value*. arXiv: 1711.04717. 2017. URL: http://arxiv.org/abs/1711.04717.

[61] J-Ph Bouchaud. "Aging in glassy systems: new experiments, simple models, and open questions". In: *Soft and Fragile Matter: Nonequilibrium Dynamics, Metastability and Flow* (2000), pp. 285–304. eprint: arXiv:cond-mat/9910387.

[62] Jean-Philippe Bouchaud. "Economics needs a scientific revolution". In: *Nature* 455.7217 (2008), pp. 1181–1181. ISSN: 1476-4687. DOI: 10.1038/4551181a. URL: https://www.nature.com/articles/4551181a.

[63] Jean-Philippe Bouchaud. “Crises and Collective Socio-Economic Phenomena: Simple Models and Challenges”. In: *Journal of Statistical Physics* 151.3-4 (2013), pp. 567–606. DOI: 10.1007/s10955-012-0687-3. URL: https://doi.org/10.1007%2Fs10955-012-0687-3.

[64] Jean-Philippe Bouchaud and Marc Mézard. “Wealth condensation in a simple model of economy”. In: *Physica A: Statistical Mechanics and its Applications* 282.3 (2000), pp. 536–545. DOI: 10.1016/S0378-4371(00)00205-3.

[65] Jean-Philippe Bouchaud and Marc Mézard. “Universality classes for extreme-value statistics”. In: *Journal of Physics A: Mathematical and General* 30.23 (1997), pp. 7997–8015. ISSN: 0305-4470, 1361-6447. DOI: 10.1088/0305-4470/30/23/004.

[66] Jean-Philippe Bouchaud and Marc Potters. *Theory of financial risk and derivative pricing : from statistical physics to risk management*. Cambridge: Cambridge University Press, 2003. ISBN: 0521819164.

[67] Jean-Philippe Bouchaud, Julius Bonart, Jonathan Donier, and Martin Gould. *Trades, quotes and prices : financial markets under the microscope*. New York: Cambridge University Press, 2018. ISBN: 110715605X.

[68] Denis Boyer, Andrea Falcón-Cortés, Luca Giuggioli, and Satya N Majumdar. “Anderson-like localization transition of random walks with resetting”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2019.5 (2019), p. 053204. DOI: 10.1088/1742-5468/ab16c2. URL: https://doi.org/10.1088%2F1742-5468%2Fab16c2.

[69] William A. Brock and Steven N. Durlauf. “Discrete Choice with Social Interactions”. In: *The Review of Economic Studies* 68.2 (2001), pp. 235–260. ISSN: 00346527, 1467937X. URL: http://www.jstor.org/stable/2695928.

[70] Bernard Bru, François Jongmans, and Eugene Seneta. “I.J. Bienaymé: Family Information and Proof of the Criticality Theorem”. In: *International Statistical Review / Revue Internationale de Statistique* 60.2 (1992), pp. 177–183. ISSN: 0306-7734. DOI: 10.2307/1403648.

[71] Mark Buchanan. “Assume nothing”. In: *Nature Physics* 5.12 (2009), pp. 855–855. ISSN: 1745-2481. DOI: 10.1038/nphys1466. URL: https://www.nature.com/articles/nphys1466.

[72] S. V. Buldyrev, L. A. N. Amaral, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, H. E. Stanley, and M. H. R. Stanley. “Scaling behavior in economics: II. Modeling of company growth”. In: *Journal de Physique I* 7.4 (1997). arXiv: cond-mat/9702085, pp. 635–650. ISSN: 1155-4304, 1286-4862. DOI: 10.1051/jp1:1997181.

[73] Philippe Bully. “Zipf, créateur de la linguistique statistique”. In: *Communication & Langages* 2.1 (1969), pp. 23–28. DOI: 10.3406/colan.1969.3726.

[74] Guy Bunin. “Interaction patterns and diversity in assembled ecological communities”. In: (2016). eprint: arXiv:1607.04734.

[75] Guy Bunin. “Ecological communities with Lotka-Volterra dynamics”. In: *Phys. Rev. E* 95 (4 2017), p. 042414. DOI: 10.1103/PhysRevE.95.042414. URL: https://link.aps.org/doi/10.1103/PhysRevE.95.042414.

[76] *CORGIS Dataset Project: billionaires CSV file*. https://corgis-edu.github.io/corgis/csv/billionaires/. Accessed: 2020-05-27.

[77] José L. Calvo. “Testing Gibrat’s Law for Small, Young and Innovating Firms”. In: *Small Business Economics* 26.2 (2006), pp. 117–123. ISSN: 1573-0913. DOI: 10.1007/s11187-004-2135-5.

[78] John Y. Campbell and John H. Cochrane. “By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior”. In: *Journal of Political Economy* 107.2 (1999), pp. 205–251. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/10.1086/250059.

[79] Christopher D. Carroll, Jody Overland, and David N. Weil. “Saving and Growth with Habit Formation”. In: *American Economic Review* 90.3 (2000), pp. 341–355. DOI: 10.1257/aer.90.3.341. URL: http://www.aeaweb.org/articles?id=10.1257/aer.90.3.341.

[80] Claudio Castellano and Romualdo Pastor-Satorras. “Relating Topological Determinants of Complex Networks to Their Spectral Properties: Structural and Dynamical Effects”. In: *Phys. Rev. X* 7 (4 2017), p. 041024. DOI: 10.1103/PhysRevX.7.041024. URL: https://link.aps.org/doi/10.1103/PhysRevX.7.041024.

[81] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos. “Epidemic thresholds in real networks”. In: *ACM Transactions on Information and System Security* 10.4 (2008), pp. 1–26. DOI: 10.1145/1284680.1284681. URL: https://doi.org/10.1145%2F1284680.1284681.

[82] D. G. Champernowne. “A Model of Income Distribution”. In: *The Economic Journal* 63.250 (1953), p. 318. ISSN: 00130133. DOI: 10.2307/2227127. URL: https://academic.oup.com/ej/article/63/250/318-351/5258723.

[83] David Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, 1987. ISBN: 9780195042771. URL: https://www.xarg.org/ref/a/0195042778/.

[84] Patrick Charbonneau, Jorge Kurchan, Giorgio Parisi, Pierfrancesco Urbani, and Francesco Zamponi. “Fractal free energy landscapes in structural glasses”. In: *Nature Communications* 5.1 (2014). DOI: 10.1038/ncomms4725. URL: https://doi.org/10.1038%2Fncomms4725.

[85] Kan Chen and Per Bak. *Self-Organized Criticality*. DOI: 10.1038/scientificamerican0191-46. URL: https://www.scientificamerican.com/article/self-organized-criticality/.

[86] Carl Chiarella. “The dynamics of speculative behaviour”. In: *Annals of Operations Research* 37.1 (1992), pp. 101–123. ISSN: 1572-9338. DOI: 10.1007/BF02071051.

[87] Luca Cipelletti, S. Manley, R. C. Ball, and D. A. Weitz. “Universal Aging Features in the Restructuring of Fractal Colloidal Gels”. In: *Phys. Rev. Lett.* 84 (10 2000), pp. 2275–2278. DOI: 10.1103/PhysRevLett.84.2275. URL: https://link.aps.org/doi/10.1103/PhysRevLett.84.2275.

[88] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. “Power-law distributions in empirical data”. In: *SIAM Review* 51.4 (2009). arXiv: 0706.1062, pp. 661–703. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/070710111.

[89] Claude Cohen-Tannoudji, Bernard Diu, and Frank Laloe. *Quantum Mechanics*. Wiley, 1991. ISBN: 978-0-471-16433-3.

[90] Célian Colon and Jean-Philippe Bouchaud. “Transition from plasticity to instability in the structure of economic networks”. In preparation.

[91] D. Considine, S. Redner, and H. Takayasu. “Comment on "Noise-induced bistability in a Monte Carlo surface-reaction model"”. In: *Physical Review Letters* 63.26 (1989), pp. 2857–2857. DOI: 10.1103/PhysRevLett.63.2857.

[92] George M. Constantinides. “Habit Formation: A Resolution of the Equity Premium Puzzle”. In: *Journal of Political Economy* 98.3 (1990), pp. 519–543. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/2937698.

[93] R. Cont. “Empirical properties of asset returns: stylized facts and statistical issues”. In: *Quantitative Finance* 1.2 (Feb. 2001), pp. 223–236. DOI: 10.1080/713665670. URL: https://doi.org/10.1080/713665670.

[94] Bernat Corominas-Murtra, Rudolf Hanel, and Stefan Thurner. “Understanding scaling through history-dependent processes with collapsing sample space”. In: *Proceedings of the National Academy of Sciences* 112.17 (2015), pp. 5348–5353. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1420946112.

[95] Codina Cotar and Vlada Limic. “Attraction time for strongly reinforced walks”. In: *The Annals of Applied Probability* 19.5 (Oct. 2009), pp. 1972–2007. DOI: 10.1214/08-aap564. URL: https://doi.org/10.1214/08-aap564.

[96] Andrea De Martino, Matteo Marsili, and Isaac Pérez Castillo. “Typical properties of large random economies with linear activities”. In: *Macroeconomic Dynamics* 11.S1 (2007), pp. 34–61. DOI: 10.1017/S1365100507060191.

[97] Pablo G. Debenedetti and Frank H. Stillinger. “Supercooled liquids and the glass transition”. In: *Nature* 410.6825 (Mar. 2001), pp. 259–267. DOI: 10.1038/35065704. URL: https://doi.org/10.1038/35065704.

[98] J. L. Deneubourg, S. Aron, S. Goss, and J. M. Pasteels. “The self-organizing exploratory pattern of the argentine ant”. In: *Journal of Insect Behavior* 3.2 (Mar. 1990), pp. 159–168. DOI: 10.1007/bf01417909. URL: https://doi.org/10.1007/bf01417909.

[99] B. Derrida. “Random-Energy Model: Limit of a Family of Disordered Models”. In: *Physical Review Letters* 45.2 (1980), pp. 79–82. DOI: 10.1103/PhysRevLett.45.79.

[100] Bernard Derrida. “Random-energy model: An exactly solvable model of disordered systems”. In: *Phys. Rev. B* 24 (5 1981), pp. 2613–2626. DOI: 10.1103/PhysRevB.24.2613. URL: https://link.aps.org/doi/10.1103/PhysRevB.24.2613.

[101] Théo Dessertaine, José Moran, Michael Benzaquen, and Jean-Philippe Bouchaud. in preparation.

[102] S.N. Dorogovtsev and J.F.F. Mendes. *Evolution of Networks*. Oxford University Press, Jan. 2003. DOI: 10.1093/acprof:oso/9780198515906.001.0001. URL: https://doi.org/10.1093/acprof:oso/9780198515906.001.0001.

[103] P. Le Doussal, M. Müller, and K. J. Wiese. “Avalanches in mean-field models and the Barkhausen noise in spin-glasses”. In: *EPL (Europhysics Letters)* 91.5 (2010), p. 57004. DOI: 10.1209/0295-5075/91/57004. URL: https://doi.org/10.1209%2F0295-5075%2F91%2F57004.

[104] Steven N Durlauf and H Peyton Young. *Social dynamics*. Vol. 4. Mit Press, 2004.

[105] Freeman J. Dyson. “A Brownian-Motion Model for the Eigenvalues of a Random Matrix”. In: *Journal of Mathematical Physics* 3.6 (1962), pp. 1191–1198. ISSN: 0022-2488. DOI: 10.1063/1.1703862.

[106] Sheer El-Showk, Miguel F. Paulos, David Poland, Slava Rychkov, David Simmons-Duffin, and Alessandro Vichi. “Solving the 3d Ising Model with the Conformal Bootstrap II. c-Minimization and Precise Critical Exponents”. In: *Journal of Statistical Physics* 157.4–5 (2014). arXiv: 1403.4545, pp. 869–914. ISSN: 0022-4715, 1572-9613. DOI: 10.1007/s10955-014-1042-7. URL: http://arxiv.org/abs/1403.4545.

[107] Paul Erdos and Alfred Renyi. “On the evolution of random graphs”. In: *Publ. Math. Inst. Hungary. Acad. Sci.* 5 (1960), pp. 17–61.

[108] Illes J. Farkas, Imre Derenyi, A.-L. Barabási, and Tamas Vicsek. “Spectra of "real-world" graphs: Beyond the semicircle law”. In: (2011). DOI: 10.1515/9781400841356.372. URL: https://doi.org/10.1515%2F9781400841356.372.

[109] J. Doyne Farmer and John Geanakoplos. *The Virtues and Vices of Equilibrium and the Future of Financial Economics*. ID 1112664. 2008. URL: https://papers.ssrn.com/abstract=1112664.

[110] Enrico Fermi. “On the Origin of the Cosmic Radiation”. In: *Physical Review* 75.8 (1949), pp. 1169–1174. ISSN: 0031-899X. DOI: 10.1103/PhysRev.75.1169.

[111] Kristen Fichthorn, Erdogan Gulari, and Robert Ziff. “Noise-induced bistability in a Monte Carlo surface-reaction model”. In: *Physical Review Letters* 63.14 (1989), pp. 1527–1530. DOI: 10.1103/PhysRevLett.63.1527.

[112] Miroslav Fiedler and Vlastimil Ptak. “On matrices with non-positive off-diagonal elements and positive principal minors”. In: *Czechoslovak Mathematical Journal* 12.3 (1962), pp. 382–400. URL: http://eudml.org/doc/12135.

[113] Dennis Fisher. *Japan Disaster Shakes Up Supply-Chain Strategies*. checked: 24.01.2019. 2011. URL: https://hbswk.hbs.edu/item/japan-disaster-shakes-up-supply-chain-strategies.

[114] Andrew T. Foerster, Pierre-Daniel G. Sarte, and Mark W. Watson. “Sectoral vs. Aggregate Shocks: A Structural Factor Analysis of Industrial Production”. In: *SSRN Electronic Journal* (2008). DOI: 10.2139/ssrn.2187904. URL: https://doi.org/10.2139%2Fssrn.2187904.

[115] Dongfeng Fu, Fabio Pammolli, S. V. Buldyrev, Massimo Riccaboni, Kaushik Matia, Kazuko Yamasaki, and H. Eugene Stanley. “The growth of business firms: Theoretical framework and empirical evidence”. In: *Proceedings of the National Academy of Sciences* 102.52 (2005), pp. 18801–18806. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0509543102.

[116] Jeffrey C. Fuhrer. “Habit Formation in Consumption and Its Implications for Monetary-Policy Models”. In: *American Economic Review* 90.3 (2000), pp. 367–390. DOI: 10.1257/aer.90.3.367. URL: http://www.aeaweb.org/articles?id=10.1257/aer.90.3.367.

[117] Yan V. Fyodorov and Boris A. Khoruzhenko. “Nonlinear analogue of the May-Wigner instability transition”. In: *Proceedings of the National Academy of Sciences* 113.25 (2016), pp. 6827–6832. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1601136113. URL: https://www.pnas.org/content/113/25/6827.

[118] Yan Fyodorov and Alexander Mirlin. “Localization in ensemble of sparse random matrices”. In: *Physical Review Letters* 67.15 (1991), pp. 2049–2052. DOI: 10.1103/physrevlett.67.2049. URL: https://doi.org/10.1103%2Fphysrevlett.67.2049.

[119] Xavier Gabaix. “Power laws in economics and finance”. In: *Annu. Rev. Econ.* 1.1 (2009), pp. 255–294. DOI: 10.1146/annurev.economics.050708.142940.

[120] Xavier Gabaix. “The Granular Origins of Aggregate Fluctuations”. In: *Econometrica* 79.3 (2011), pp. 733–772. DOI: 10.3982/ecta8769. URL: https://doi.org/10.3982%2Fecta8769.

[121] Xavier Gabaix, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. “The Dynamics of Inequality”. In: *Econometrica* 84.6 (2016), pp. 2071–2111. ISSN: 0012-9682. DOI: 10.3982/ECTA13569.

[122] Serge Galam, Yuval Gefen (Feigenblat), and Yonathan Shapir. “Sociophysics: A new approach of sociological collective behaviour: I. Mean-behaviour description of a strike”. In: *Journal of Mathematical Sociology* 9.1 (1982), pp. 1–13. ISSN: 0022-250X(Print). DOI: 10.1080/0022250X.1982.9989929.

[123] Jordi Galí. *Monetary policy, inflation, and the business cycle : an introduction to the new Keynesian framework and its applications*. Princeton Oxford: Princeton University Press, 2015. ISBN: 9780691164786.

[124] Mauro Gallegati and Domenico Mignacca. “Jevons, sunspot theory and economic fluctuations”. In: *History of Economic Ideas* 2.2 (1994), pp. 23–40. ISSN: 11228792, 17242169. URL: http://www.jstor.org/stable/23722216.

[125] Mauro Gallegati, Gianfranco Giulioni, Alan Kirman, and Antonio Palestrini. “What’s that got to do with the price of fish? Buyers behavior on the Ancona fish market”. In: *Journal of Economic Behavior & Organization*. Special Section: Fish Markets 80.1 (2011), pp. 20–33. ISSN: 0167-2681. DOI: 10.1016/j.jebo.2011.01.011. URL: http://www.sciencedirect.com/science/article/pii/S0167268111000278.

[126] Mark R. Gardner and W. Ross Ashby. "Connectance of Large Dynamic (Cybernetic) Systems: Critical Values for Stability". In: *Nature* 228.5273 (1970), pp. 784–784. ISSN: 1476-4687. DOI: 10.1038/228784a0. URL: https://www.nature.com/articles/228784a0.

[127] Stratis Gavaris. "Use of a multiplicative model to estimate catch rate and effort from commercial data". In: *Canadian Journal of Fisheries and Aquatic Sciences* 37.12 (1980), pp. 2272–2275.

[128] R. Gibrat. *Les inégalités économiques: applications: aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effect proportionnel*. Recueil Sirey, 1931. URL: https://books.google.fr/books?id=m9fuoAEACAAJ.

[129] Gerd Gigerenzer and Reinhard Selten. *Bounded rationality: The adaptive toolbox*. MIT press, 2002. ISBN: 9780262072144.

[130] M. Gilli and P. Winker. "A global optimization heuristic for estimating agent based models". In: *Computational Statistics & Data Analysis* 42.3 (Mar. 2003), pp. 299–312. DOI: 10.1016/s0167-9473(02)00214-1. URL: https://doi.org/10.1016/s0167-9473(02)00214-1.

[131] Julian di Giovanni, Andrei A. Levchenko, and Isabelle Méjean. "Large Firms and International Business Cycle Comovement". In: *American Economic Review* 107.5 (2017), pp. 598–602. DOI: 10.1257/aer.p20171006. URL: https://doi.org/10.1257%2Faer.p20171006.

[132] V. L. Girko. "Circular Law". In: *Theory of Probability & Its Applications* 29.4 (1985), pp. 694–706. DOI: 10.1137/1129095. URL: https://doi.org/10.1137%2F1129095.

[133] *Global Fishing Watch, Fishing Vessels Dataset, https://globalfishingwatch.org/datasets-and-code/vessel-identity*. 2020. URL: https://globalfishingwatch.org/datasets-and-code/vessel-identity/.

[134] P. Gopikrishnan, M. Meyer, L. a. N. Amaral, and H. E. Stanley. "Inverse cubic law for the distribution of stock price variations". In: *The European Physical Journal B* 3.2 (1998), pp. 139–140. ISSN: 1434-6028, 1434-6036. DOI: 10.1007/s100510050292.

[135] Mirta B. Gordon, Jean-Pierre Nadal, Denis Phan, and Jean Vannimenus. "Seller's dilemma due to social interactions between customers". In: *Physica A: Statistical Mechanics and its Applications* 356.2-4 (2005), pp. 628–640. ISSN: 03784371. DOI: 10.1016/j.physa.2005.03.003.

[136] Stanislao Gualdi and Antoine Mandel. "On the emergence of scale-free production networks". In: *arXiv:1509.01483 [physics, q-fin]* (2016). arXiv: 1509.01483. URL: http://arxiv.org/abs/1509.01483.

[137] Marshall Hall and Nicolaus Tideman. "Measures of Concentration". In: *Journal of the American Statistical Association* 62.317 (Mar. 1967), pp. 162–168. DOI: 10.1080/01621459.1967.10482897. URL: https://doi.org/10.1080/01621459.1967.10482897.

[138] Rudolf Hanel and Stefan Thurner. "The role of grammar in transition-probabilities of subsequent words in English text". In: *arXiv:1812.10991 [physics]* (2018). arXiv: 1812.10991. URL: http://arxiv.org/abs/1812.10991.

[139] Garrett Hardin. "The tragedy of the commons". In: *Science* 162.3859 (1968), pp. 1243–1248.

[140] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (Apr. 1970), pp. 97–109. DOI: 10.1093/biomet/57.1.97. URL: https://doi.org/10.1093/biomet/57.1.97.

[141] Naomichi Hatano and David R. Nelson. "Localization Transitions in Non-Hermitian Quantum Mechanics". In: *Physical Review Letters* 77.3 (1996), pp. 570–573. DOI: 10.1103/physrevlett.77.570. URL: https://doi.org/10.1103%2Fphysrevlett.77.570.

[142] David Hawkins. "Some Conditions of Macroeconomic Stability". In: *Econometrica* 16.4 (1948), p. 309. DOI: 10.2307/1909272. URL: https://doi.org/10.2307%2F1909272.

[143] David Hawkins and Herbert A. Simon. "Note: Some Conditions of Macroeconomic Stability". In: *Econometrica* 17.3/4 (1949), p. 245. DOI: 10.2307/1905526. URL: https://doi.org/10.2307%2F1905526.

[144] Takashi Ichinomiya. "Bouchaud-Mézard model on a random network". In: *Physical Review E* 86.3 (2012). arXiv: 1209.2467, p. 036111. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.86.036111.

[145] Kiyosi Ito. *On stochastic differential equations*. 4. American Mathematical Society (AMS), 1951. DOI: 10.1090/memo/0004. URL: https://doi.org/10.1090/memo/0004.

[146] Robert L. Jack and Rosemary J. Harris. *Giant leaps and long excursions: fluctuation mechanisms in systems with long-range memory*. 2020. eprint: arXiv:2003.03587.

[147] Daniel Kahneman. "Experienced Utility and Objective Happiness: A Moment-Based Approach". In: *Choices, Values, and Frames*. Cambridge University Press, Sept. 2000, pp. 673–692. DOI: 10.1017/cbo9780511803475.038. URL: https://doi.org/10.1017/cbo9780511803475.038.

[148] Daniel Kahneman. “A Psychological Perspective on Economics”. In: *American Economic Review* 93.2 (Apr. 2003), pp. 162–168. DOI: 10.1257/000282803321946985. URL: https://doi.org/10.1257/000282803321946985.

[149] Daniel Kahneman and Richard H. Thaler. “Anomalies: Utility Maximization and Experienced Utility”. In: *Journal of Economic Perspectives* 20.1 (2006), pp. 221–234. DOI: 10.1257/089533006776526076. URL: http://www.aeaweb.org/articles?id=10.1257/089533006776526076.

[150] Mehran Kardar. *Statistical physics of fields*. OCLC: ocn123113789. Cambridge ; New York: Cambridge University Press, 2007. ISBN: 9780521873413.

[151] Harry Kesten. “Random difference equations and Renewal theory for products of random matrices”. In: *Acta Mathematica* 131 (1973). Zbl: 0291.60029, pp. 207–248. ISSN: 0001-5962, 1871-2509. DOI: 10.1007/BF02392040.

[152] Jonathan Khedair and Reimer Kühn. “A structural model of market dynamics, and why it matters”. In: *Comptes Rendus Physique* 20.4 (2019), pp. 336–348. ISSN: 1631-0705. DOI: 10.1016/j.crhy.2019.05.013. URL: http://www.sciencedirect.com/science/article/pii/S1631070519300428.

[153] M. Kimura. “Stochastic processes and distribution of gene frequencies under natural selection”. In: *Cold Spring Harbor Symposia on Quantitative Biology* 20.0 (Jan. 1955), pp. 33–53. DOI: 10.1101/sqb.1955.020.01.006. URL: https://doi.org/10.1101/sqb.1955.020.01.006.

[154] Alan P Kirman. “Whom or What Does the Representative Individual Represent?” In: *Journal of Economic Perspectives* 6.2 (1992), pp. 117–136. ISSN: 0895-3309. DOI: 10.1257/jep.6.2.117. URL: https://pubs.aeaweb.org/doi/10.1257/jep.6.2.117.

[155] Alan P. Kirman and Nicolaas J. Vriend. “Learning to Be Loyal. A Study of the Marseille Fish Market”. In: *Interaction and Market Structure*. Ed. by Domenico Delli Gatti, Mauro Gallegati, and Alan Kirman. Lecture Notes in Economics and Mathematical Systems. Springer, 2000, pp. 33–56. ISBN: 978-3-642-57005-6. DOI: 10.1007/978-3-642-57005-6_3.

[156] Alan P. Kirman and Nicolaas J. Vriend. “Evolving market structure: An ACE model of price dispersion and loyalty”. In: *Journal of Economic Dynamics and Control* 25.3-4 (Mar. 2001), pp. 459–502. DOI: 10.1016/s0165-1889(00)00033-6. URL: https://doi.org/10.1016/s0165-1889(00)00033-6.

[157] Alan Kirman. “Ants, Rationality, and Recruitment”. In: *The Quarterly Journal of Economics* 108.1 (1993), pp. 137–156. ISSN: 0033-5533. DOI: 10.2307/2118498.

[158] Alan Kirman. "The Complex Nature of Economic Liberalism". In: *History of economic ideas* XXIV (2016). ISSN: 1724-2169. DOI: 10.19272/201606103003. URL: http://doi.org/10.19272/201606103003.

[159] Alan Kirman and Teyssière Gilles. "Microeconomic Models for Long Memory in the Volatility of Financial Time Series". In: *Studies in Nonlinear Dynamics & Econometrics* 5.4 (2002), pp. 1–23. URL: https://EconPapers.repec.org/RePEc:bpj:sndecm:v:5:y:2002:i:4:n:3.

[160] P. L. Krapivsky and S. Redner. "Organization of Growing Random Networks". In: *Physical Review E* 63.6 (2001). arXiv: cond-mat/0011094, p. 066123. ISSN: 1063-651X, 1095-3787. DOI: 10.1103/PhysRevE.63.066123.

[161] Reimer Kühn. "Spectra of sparse random matrices". In: *Journal of Physics A: Mathematical and Theoretical* 41.29 (2008), p. 295002. DOI: 10.1088/1751-8113/41/29/295002. URL: https://doi.org/10.1088%2F1751-8113%2F41%2F29%2F295002.

[162] Shin-Ichiro Kumamoto and Takashi Kamihigashi. "Power Laws in Stochastic Processes for Social Phenomena: An Introductory Review". In: *Frontiers in Physics* 6 (2018). ISSN: 2296-424X. DOI: 10.3389/fphy.2018.00020. URL: https://www.frontiersin.org/articles/10.3389/fphy.2018.00020/full.

[163] R. Lambiotte and S. Redner. "Dynamics of Vacillating Voters". In: *Journal of Statistical Mechanics: Theory and Experiment* 2007.10 (2007). arXiv: 0710.0914, pp. L10001–L10001. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2007/10/L10001.

[164] Youngki Lee, Luis A. N. Amaral, David Canning, Martin Meyer, and H. Eugene Stanley. "Universal features in the growth dynamics of complex organizations". In: *Physical Review Letters* 81.15 (1998). arXiv: cond-mat/9804100, pp. 3275–3278. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.81.3275.

[165] Wassily W. Leontief. "Quantitative Input and Output Relations in the Economic Systems of the United States". In: *The Review of Economics and Statistics* 18.3 (Aug. 1936), p. 105. DOI: 10.2307/1927837. URL: https://doi.org/10.2307/1927837.

[166] W. Li. "Random texts exhibit Zipf's-law-like word frequency distribution". In: *IEEE Transactions on Information Theory* 38.6 (1992), pp. 1842–1845. ISSN: 1557-9654. DOI: 10.1109/18.165464.

[167] I.M. Lifshitz. "The energy spectrum of disordered systems". In: *Advances in Physics* 13.52 (1964), pp. 483–536. DOI: 10.1080/00018736400101061. URL: https://doi.org/10.1080%2F00018736400101061.

[168] Giacomo Livan, Marcel Novaes, and Pierpaolo Vivo. "Introduction to Random Matrices - Theory and Practice". In: *arXiv:1712.07903 [cond-mat, physics:math-ph]* 26 (2018). arXiv: 1712.07903. DOI: 10.1007/978-3-319-70885-0. URL: http://arxiv.org/abs/1712.07903.

[169] John B. Long and Charles I. Plosser. "Real Business Cycles". In: *Journal of Political Economy* 91.1 (1983), pp. 39–69. DOI: 10.1086/261128. URL: https://doi.org/10.1086%2F261128.

[170] Francesca Lotti, Enrico Santarelli, and Marco Vivarelli. "Does Gibrat's Law hold among young, small firms?" In: *Journal of Evolutionary Economics* 13.3 (2003), pp. 213–235. ISSN: 1432-1386. DOI: 10.1007/s00191-003-0153-0.

[171] Thomas Lux. "Herd Behaviour, Bubbles and Crashes". In: *The Economic Journal* 105.431 (July 1995), p. 881. DOI: 10.2307/2235156. URL: https://doi.org/10.2307/2235156.

[172] Robert Mac Arthur. "Species packing, and what competition minimizes". In: *Proceedings of the National Academy of Sciences* 64.4 (1969), pp. 1369–1371.

[173] Robert MacArthur. "Species packing and competitive equilibrium for many species". In: *Theoretical population biology* 1.1 (1970), pp. 1–11.

[174] Adam A. Majewski, Stefano Ciliberti, and Jean-Philippe Bouchaud. "Co-existence of trend and value in financial markets: Estimating an extended Chiarella model". In: *Journal of Economic Dynamics and Control* 112 (2020), p. 103791. ISSN: 0165-1889. DOI: 10.1016/j.jedc.2019.103791.

[175] Antoine Mandel and Vipin P. Veetil. *The Economic Cost of COVID Lockdowns: An Out-of-Equilibrium Analysis*. ID 3588421. 2020. URL: https://papers.ssrn.com/abstract=3588421.

[176] Benoit Mandelbrot. "An informational theory of the statistical structure of language". In: *Communication theory* 84 (1953), pp. 486–502.

[177] Benoit Mandelbrot. "Information theory and psycholinguistics". In: *BB Wolman and E* (1965).

[178] V. A. Marčenko and L. A. Pastur. "Distribution of eigenvalues for some sets of random matrices". In: *Mathematics of the USSR-Sbornik* 1.4 (1967), p. 457. ISSN: 0025-5734. DOI: 10.1070/SM1967v001n04ABEH001994.

[179] Robert M. May. "Will a Large Complex System be Stable?" In: *Nature* 238.5364 (1972), pp. 413–414. DOI: 10.1038/238413a0. URL: https://doi.org/10.1038%2F238413a0.

[180] A.J. McKane and D. Waxman. “Singular solutions of the diffusion equation of population genetics”. In: *Journal of Theoretical Biology* 247.4 (Aug. 2007), pp. 849–858. DOI: 10.1016/j.jtbi.2007.04.016. URL: https://doi.org/10.1016/j.jtbi.2007.04.016.

[181] Brendan D. McKay. “The expected eigenvalue distribution of a large regular graph”. In: *Linear Algebra and its Applications* 40 (1981), pp. 203–216. DOI: 10.1016/0024-3795(81)90150-6. URL: https://doi.org/10.1016%2F0024-3795%2881%2990150-6.

[182] Nicholas Metropolis and S. Ulam. “The Monte Carlo Method”. In: *Journal of the American Statistical Association* 44.247 (Sept. 1949), pp. 335–341. DOI: 10.1080/01621459.1949.10483310. URL: https://doi.org/10.1080/01621459.1949.10483310.

[183] Fernando Lucas Metz, Izaak Neri, and Tim Rogers. “Spectra of Sparse Non-Hermitian Random Matrices”. In: (2018). eprint: arXiv:1811.10416.

[184] Fernando Lucas Metz, Izaak Neri, and Tim Rogers. “Spectral theory of sparse non-Hermitian random matrices”. In: *Journal of Physics A: Mathematical and Theoretical* 52.43 (2019), p. 434003. DOI: 10.1088/1751-8121/ab1ce0. URL: https://doi.org/10.1088%2F1751-8121%2Fab1ce0.

[185] Quentin Michard and Jean-Philippe Bouchaud. “Theory of collective opinion shifts: from smooth trends to abrupt swings”. In: *The European Physical Journal B* 47.1 (2005). arXiv: cond-mat/0504079, pp. 151–159. ISSN: 1434-6028, 1434-6036. DOI: 10.1140/epjb/e2005-00307-0.

[186] George A. Miller. “Some Effects of Intermittent Silence”. In: *The American Journal of Psychology* 70.2 (1957), pp. 311–314. ISSN: 0002-9556. DOI: 10.2307/1419346.

[187] R. Monasson. “Diffusion, localization and dispersion relations on “small-world” lattices”. In: *The European Physical Journal B* 12.4 (1999), pp. 555–567. DOI: 10.1007/s100510051038. URL: https://doi.org/10.1007%2Fs100510051038.

[188] Cécile Monthus and Jean-Philippe Bouchaud. “Models of traps and glass phenomenology”. In: *Journal of Physics A: Mathematical and General* 29.14 (1996), pp. 3847–3869. DOI: 10.1088/0305-4470/29/14/012. URL: https://doi.org/10.1088%2F0305-4470%2F29%2F14%2F012.

[189] José Moran, Antoine Fosset, Michael Benzaquen, and Jean-Philippe Bouchaud. “Schrödinger’s ants: A continuous description of Kirman’s recruitment model”. In: *Journal of Physics: Complexity* (June 2020). DOI: 10.1088/2632-072x/aba115. URL: https://doi.org/10.1088/2632-072x/aba115.

[190] José Moran and Jean-Philippe Bouchaud. "Greedy algorithms and Zipf laws". In: *Journal of Statistical Mechanics: Theory and Experiment* 2018.4 (2018). arXiv: 1801.05279, p. 043402. ISSN: 1742-5468. DOI: 10.1088/1742-5468/aab50a.

[191] José Moran and Jean-Philippe Bouchaud. "May's Instability in Large Economies". In: *Physical Review E* 100.3 (2019). arXiv: 1901.09629, p. 032307. ISSN: 2470-0045, 2470-0053. DOI: 10.1103/PhysRevE.100.032307.

[192] José Moran, Antoine Fosset, Davide Luzzati, Jean-Philippe Bouchaud, and Michael Benzaquen. "By force of habit: Self-trapping in a dynamical utility landscape". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30.5 (2020), p. 053123. ISSN: 1054-1500. DOI: 10.1063/5.0009518.

[193] P. A. P. Moran. "Random processes in genetics". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 54.1 (1958), pp. 60–71. DOI: 10.1017/S0305004100033193.

[194] Federico Guglielmo Morelli, Michael Benzaquen, Marco Tarzia, and Jean-Philippe Bouchaud. "Confidence collapse in a multihousehold, self-reflexive DSGE model". In: *Proceedings of the National Academy of Sciences* 117.17 (2020), pp. 9244–9249. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1912280117. URL: https://www.pnas.org/content/117/17/9244.

[195] Markus Müller and Matthieu Wyart. "Marginal Stability in Structural, Spin, and Electron Glasses". In: *Annual Review of Condensed Matter Physics* 6.1 (2015), pp. 177–200. DOI: 10.1146/annurev-conmatphys-031214-014614. URL: https://doi.org/10.1146%2Fannurev-conmatphys-031214-014614.

[196] Jean-Pierre Nadal, Denis Phan, Mirta B. Gordon, and Jean Vannimenus. "Monopoly Market with Externality: an Analysis with Statistical Physics and Agent Based Computational Economics". In: *arXiv:cond-mat/0311096* (2003). arXiv: cond-mat/0311096. URL: http://arxiv.org/abs/cond-mat/0311096.

[197] David R. Nelson and Nadav M. Shnerb. "Non-Hermitian localization and population biology". In: *Physical Review E* 58.2 (1998), pp. 1383–1403. DOI: 10.1103/physreve.58.1383. URL: https://doi.org/10.1103%2Fphysreve.58.1383.

[198] Izaak Neri and Fernando Lucas Metz. "Eigenvalue Outliers of Non-Hermitian Random Matrices with a Local Tree Structure". In: *Phys. Rev. Lett.* 117 (22 2016), p. 224101. DOI: 10.1103/PhysRevLett.117.224101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.117.224101.

[199] Mark Newman. *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010. ISBN: 0199206651, 9780199206650.

[200] Edward W. Ng and Murray Geller. "A table of integrals of the Error functions". In: *Journal of Research of the National Bureau of Standards, Section B: Mathematical Sciences* 73B.1 (1969), p. 1. ISSN: 0098-8979. DOI: 10.6028/jres.073B.001.

[201] Michael Martin Nieto and L. M. Simmons. "Coherent states for general potentials. II. Confining one-dimensional examples". In: *Phys. Rev. D* 20 (6 1979), pp. 1332–1341. DOI: 10.1103/PhysRevD.20.1332. URL: https://link.aps.org/doi/10.1103/PhysRevD.20.1332.

[202] Makoto Nirei and José Scheinkman. "Self-Organization of Inflation Volatility". In: *SSRN Electronic Journal* (2019). ISSN: 1556-5068. DOI: 10.2139/ssrn.3396528. URL: https://www.ssrn.com/abstract=3396528.

[203] *OECD Economic Surveys: Finland 2004*. OECD, Oct. 2004. DOI: 10.1787/eco_surveys-fin-2004-en. URL: https://doi.org/10.1787/eco_surveys-fin-2004-en.

[204] Norimitsu Onishi and Constant Méheut. "How France Lost the Weapons to Fight a Pandemic". In: *The New York Times* (2020). ISSN: 0362-4331. URL: https://www.nytimes.com/2020/05/17/world/europe/france-coronavirus.html.

[205] Lars Onsager. "Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition". In: *Physical Review* 65.3–4 (1944), pp. 117–149. ISSN: 0031-899X. DOI: 10.1103/PhysRev.65.117. URL: https://link.aps.org/doi/10.1103/PhysRev.65.117.

[206] Richard Otter. "The Multiplicative Process". In: *Annals of Mathematical Statistics* 20.2 (1949). Zbl: 0033.38301, pp. 206–224. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177730031.

[207] Marco Pangallo. "Synchronization of endogenous business cycles". In: *arXiv:2002.06555 [nlin, q-fin]* (2020). arXiv: 2002.06555. URL: http://arxiv.org/abs/2002.06555.

[208] Vilfredo Pareto. *Cours d'économie politique*. Librairie Droz, 1964. ISBN: 978-2-600-04014-3. DOI: 10.3917/droz.paret.1964.01. URL: https://www.cairn.info/cours-d-economie-politique-tomes-1-et-2--9782600040143.htm.

[209] Robin Pemantle. "A survey of random processes with reinforcement". In: *Probab. Surveys* 4 (2007), pp. 1–79. DOI: 10.1214/07-PS094. URL: https://doi.org/10.1214/07-PS094.

[210] Richard Perline and Ronald Perline. "Two Universality Properties Associated with the Monkey Model of Zipf's Law". In: *Entropy* 18.3 (2016). arXiv: 1511.09449, p. 89. ISSN: 1099-4300. DOI: 10.3390/e18030089.

[211] Juan I. Perotti, Orlando V. Billoni, Francisco A. Tamarit, Dante R. Chialvo, and Sergio A. Cannas. "Emergent Self-Organized Complex Network Topology out of Stability Constraints". In: *Phys. Rev. Lett.* 103 (10 2009), p. 108701. DOI: 10.1103/PhysRevLett.103.108701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.108701.

[212] Ole Peters. "The ergodicity problem in economics". In: *Nature Physics* 15.12 (2019), pp. 1216–1221. ISSN: 1745-2481. DOI: 10.1038/s41567-019-0732-0. URL: https://www.nature.com/articles/s41567-019-0732-0.

[213] Anton Pichler, Marco Pangallo, R. Maria del Río-Chanona, François Lafond, and J. Doyne Farmer. "Production Networks and Epidemic Spreading: How to Restart the UK Economy?" In: *SSRN Electronic Journal* (2020). ISSN: 1556-5068. DOI: 10.2139/ssrn.3606984. URL: https://www.ssrn.com/abstract=3606984.

[214] Thomas Piketty. *Le capital au XXIe siècle*. Paris: Éditions du Seuil, 2013. ISBN: 978-2-02-108228-9.

[215] Thomas Piketty. *Capital et idéologie*. Paris: Seuil, 2019. ISBN: 9782021338065.

[216] Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. "Distributional National Accounts: Methods and Estimates for the United States". In: *The Quarterly Journal of Economics* 133.2 (2018), pp. 553–609. ISSN: 0033-5533. DOI: 10.1093/qje/qjx043.

[217] R.J. Plemmons. "M-matrix characterizations.I—nonsingular M-matrices". In: *Linear Algebra and its Applications* 18.2 (1977), pp. 175–188. DOI: 10.1016/0024-3795(77)90073-8. URL: https://doi.org/10.1016%2F0024-3795%2877%2990073-8.

[218] Vasiliki Plerou, Parameswaran Gopikrishnan, Luís A. Nunes Amaral, Martin Meyer, and H. Eugene Stanley. "Scaling of the distribution of price fluctuations of individual companies". In: *Physical Review E* 60.6 (Dec. 1999), pp. 6519–6529. DOI: 10.1103/physreve.60.6519. URL: https://doi.org/10.1103/physreve.60.6519.

[219] Robert A. Pollak. "Habit Formation and Dynamic Demand Functions". In: *Journal of Political Economy* 78.4 (1970), pp. 745–763. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/1829929.

[220] George Poole and Thomas Boullion. "A survey on M-matrices". In: *SIAM review* 16.4 (1974), pp. 419–427. DOI: 10.1137/1016079.

[221] Theodore M. Porter. "A Statistical Survey of Gases: Maxwell's Social Physics". In: *Historical Studies in the Physical Sciences* 12.1 (1981), pp. 77–116. ISSN: 0073-2672. DOI: 10.2307/27757490. URL: https://www.jstor.org/stable/27757490.

[222] Marc Potters and Jean-Philippe Bouchaud. “A First Course in Random Matrix Theory”. in preparation. URL: https://physics-complex-systems.fr/wp-content/uploads/2019/02/Notes_chap1-11.pdf.

[223] François Quesnay. “Analyse de la formule arithmétique du tableau économique de la distribution des dépenses annuelles d’une nation agricole”. In: *Journal d’agriculture, du commerce et des finances* 5.3 (June 1766), pp. 11–41.

[224] Devesh R. Raval. “The micro elasticity of substitution and non-neutral technology”. In: *The RAND Journal of Economics* (2019). DOI: 10.1111/1756-2171.12265. URL: https://doi.org/10.1111%2F1756-2171.12265.

[225] John Reed and Bernard Simon. *Car components hit by Japan aftershock*. checked: 24.01.2019. 2011. URL: https://www.ft.com/content/9ac4d7e2-595f-11e0-bc39-00144feab49a.

[226] Hannes Risken. “Fokker-Planck equation”. In: *The Fokker-Planck Equation*. Springer Berlin Heidelberg, 1996, pp. 63–95.

[227] Tim Rogers and Isaac Pérez Castillo. “Cavity approach to the spectral density of non-Hermitian sparse matrices”. In: *Phys. Rev. E* 79 (1 2009), p. 012101. DOI: 10.1103/PhysRevE.79.012101. URL: https://link.aps.org/doi/10.1103/PhysRevE.79.012101.

[228] Felix Roy, Matthieu Barbier, Giulio Biroli, and Guy Bunin. “Can endogenous fluctuations persist in high-diversity ecosystems?” In: *arXiv:1908.03348 [cond-mat, physics:physics, q-bio]* (2019). arXiv: 1908.03348. URL: http://arxiv.org/abs/1908.03348.

[229] Alexander I Saichev, Yannick Malevergne, and Didier Sornette. *Theory of Zipf’s law and beyond*. Vol. 632. Springer Science & Business Media, 2009. DOI: 10.1007/978-3-642-02946-2_1.

[230] David S. Scharfstein and Jeremy C. Stein. “Herd Behavior and Investment”. In: *The American Economic Review* 80.3 (1990), pp. 465–479. ISSN: 00028282. URL: http://www.jstor.org/stable/2006678.

[231] José Scheinkman and Michael Woodford. “Self-Organized Criticality and Economic Fluctuations”. In: *American Economic Review* 84.2 (1994), pp. 417–21. URL: https://EconPapers.repec.org/RePEc:aea:aecrev:v:84:y:1994:i:2:p:417-21.

[232] Thomas C. Schelling. *Micromotives and Macrobehavior*. Oct. 2006. ISBN: 9780393329469.

[233] Reinhard Selten. “Bounded Rationality”. In: *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft* 146.4 (1990), pp. 649–658. ISSN: 09324569. URL: http://www.jstor.org/stable/40751353.

[234] J. Sethna. *Statistical Mechanics: Entropy, Order Parameters and Complexity*. Oxford Master Series in Physics. OUP Oxford, 2006. ISBN: 9780198566779.

[235] James P. Sethna, Karin A. Dahmen, and Christopher R. Myers. "Crackling noise". In: *Nature* 410.6825 (2001), pp. 242–250. ISSN: 1476-4687. DOI: 10.1038/35065675.

[236] James P. Sethna, Karin Dahmen, Sivan Kartha, James A. Krumhansl, Bruce W. Roberts, and Joel D. Shore. "Hysteresis and hierarchies: dynamics of disorder-driven first-order phase transformations". In: *Physical Review Letters* 70.21 (1993). arXiv: cond-mat/9210018, pp. 3347–3350. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.70.3347.

[237] Dhruv Sharma, Jean-Philippe Bouchaud, Marco Tarzia, and Francesco Zamponi. "A constraint-satisfaction agent-based model for the macroeconomy". In: *arXiv:2005.11748 [physics, q-fin]* (2020). arXiv: 2005.11748. URL: http://arxiv.org/abs/2005.11748.

[238] David Sherrington and Scott Kirkpatrick. "Solvable Model of a Spin-Glass". In: *Physical Review Letters* 35.26 (1975), pp. 1792–1796. DOI: 10.1103/PhysRevLett.35.1792.

[239] Robert J. Shiller and John Pound. "Survey evidence on diffusion of interest and information among investors". In: *Journal of Economic Behavior & Organization* 12.1 (Aug. 1989), pp. 47–66. DOI: 10.1016/0167-2681(89)90076-0. URL: https://doi.org/10.1016/0167-2681(89)90076-0.

[240] M. V. Simkin and V. P. Roychowdhury. "Re-inventing Willis". In: *Physics Reports* (2010). arXiv: physics/0601192, S0370157310003339. ISSN: 03701573. DOI: 10.1016/j.physrep.2010.12.004.

[241] Herbert A. Simon. "A Behavioral Model of Rational Choice". In: *The Quarterly Journal of Economics* 69.1 (1955), pp. 99–118. ISSN: 00335533, 15314650. URL: http://www.jstor.org/stable/1884852.

[242] Herbert A Simon. "On a class of skew distribution functions". In: *Biometrika* 42.3/4 (1955), pp. 425–440. DOI: 10.1093/biomet/42.3-4.425.

[243] Herbert A Simon. "Theories of bounded rationality". In: *Decision and organization* 1.1 (1972), pp. 161–176.

[244] Herbert A. Simon and Charles P. Bonini. "The Size Distribution of Business Firms". In: *The American Economic Review* 48.4 (1958), pp. 607–617. ISSN: 0002-8282. URL: https://www.jstor.org/stable/1808270.

[245] D. Sornette. "Endogenous versus Exogenous Origins of Crises". In: *arXiv:physics/0412026* (2004). arXiv: physics/0412026. URL: http://arxiv.org/abs/physics/0412026.

[246] Didier Sornette. “Dragon-Kings, Black Swans and the Prediction of Crises”. In: *arXiv:0907.4290 [physics]* (2009). arXiv: 0907.4290. URL: http://arxiv.org/abs/0907.4290.

[247] Didier Sornette and Rama Cont. “Convergent multiplicative processes repelled from zero: power laws and truncated power laws”. In: *Journal de Physique I* 7.3 (1997), pp. 431–444. DOI: 10.1051/jp1:1997169.

[248] Didier Sornette and Guy Ouillon. “Dragon-kings: mechanisms, statistical methods and empirical evidence”. In: (), p. 20.

[249] Wataru Souma, Yoshi Fujiwara, and Hideaki Aoyama. “Small-World Effects in Wealth Distribution”. In: *arXiv:cond-mat/0108482* (2001). arXiv: cond-mat/0108482. URL: http://arxiv.org/abs/cond-mat/0108482.

[250] John L. Spouge. “Polymers and random graphs: Asymptotic equivalence to branching processes”. In: *Journal of Statistical Physics* 38.3 (1985), pp. 573–587. ISSN: 1572-9613. DOI: 10.1007/BF01010478.

[251] Michael H. R. Stanley, Luís A. N. Amaral, Sergey V. Buldyrev, Shlomo Havlin, Heiko Leschhorn, Philipp Maass, Michael A. Salinger, and H. Eugene Stanley. “Scaling behaviour in the growth of companies”. In: *Nature* 379.6568 (1996), pp. 804–806. ISSN: 1476-4687. DOI: 10.1038/379804a0.

[252] Joseph E Stiglitz. “Where modern macroeconomics went wrong”. In: *Oxford Review of Economic Policy* 34.1-2 (2018), pp. 70–106. DOI: 10.1093/oxrep/grx057. URL: https://doi.org/10.1093/oxrep/grx057.

[253] Vito A. R. Susca, Pierpaolo Vivo, and Reimer Kühn. “Top eigenpair statistics for weighted sparse graphs”. In: *Journal of Physics A: Mathematical and Theoretical* 52.48 (2019), p. 485002. ISSN: 1751-8121. DOI: 10.1088/1751-8121/ab4d63. URL: https://doi.org/10.1088%2F1751-8121%2Fab4d63.

[254] John Sutton. “The variance of firm growth rates: the ‘scaling’ puzzle”. In: *Physica A: Statistical Mechanics and its Applications* 312.3-4 (2002), pp. 577–590. DOI: 10.1016/s0378-4371(02)00852-x. URL: https://doi.org/10.1016%2Fs0378-4371%2802%2900852-x.

[255] Chad Syverson. *What Determines Productivity?* Working Paper 15712. National Bureau of Economic Research, 2010. DOI: 10.3386/w15712. URL: http://www.nber.org/papers/w15712.

[256] Naomi Tajitsu. *Five years after Japan quake, rewiring of auto supply chain hits limits*. checked: 24.01.2019. 2016. URL: https://www.reuters.com/article/us-japan-quake-supplychain-idUSKCN0WW09N.

[257] Nassim Nicholas Taleb. *The Black Swan: The Impact of the Highly Improbable*. Random House Group, 2007. ISBN: 1400063515.

[258] Hidenori Tanaka and David R Nelson. “Non-Hermitian Quasi-Localization and Ring Attractor Neural Networks”. In: (2018). eprint: arXiv:1811.07433.

[259] Terence Tao and American Mathematical Society. *Topics in random matrix theory*. English. OCLC: 1132122121. 2012. ISBN: 9780821885062. URL: http://www.ams.org/gsm/132 (visited on 06/26/2020).

[260] Terence Tao and Van Vu. “Random Matrices: the circular law”. In: *Communications in Contemporary Mathematics* 10.02 (2008), pp. 261–307. DOI: 10.1142/s0219199708002788. URL: https://doi.org/10.1142%2Fs0219199708002788.

[261] H. Taşeli. “Exact Analytical Solutions of the Hamiltonian with a Squared Tangent Potential”. In: *Journal of Mathematical Chemistry* 34.3/4 (Nov. 2003), pp. 243–251. DOI: 10.1023/b:jomc.0000004073.17023.41. URL: https://doi.org/10.1023/b:jomc.0000004073.17023.41.

[262] D.J. Thouless. “Electrons in disordered systems and the theory of localization”. In: *Physics Reports* 13.3 (1974), pp. 93–142. DOI: 10.1016/0370-1573(74)90029-5. URL: https://doi.org/10.1016%2F0370-1573%2874%2990029-5.

[263] S. Thurner, R. Hanel, B. Liu, and B. Corominas-Murtra. “Understanding Zipf’s law of word frequencies through sample-space collapse in sentence formation”. In: *Journal of the Royal Society Interface* 12 (2015), p. 20150330. DOI: 10.1098/rsif.2015.0330. URL: http://rsif.royalsocietypublishing.org/content/12/108/20150330.

[264] F. Vazquez and S. Redner. “Ultimate Fate of Constrained Voters”. In: *Journal of Physics A: Mathematical and General* 37.35 (2004). arXiv: cond-mat/0405652, pp. 8479–8494. ISSN: 0305-4470, 1361-6447. DOI: 10.1088/0305-4470/37/35/006.

[265] Vincent Verbavatz and Marc Barthélemy. “The growth equation of cities”. in preparation.

[266] Marianne Vignaux. “Analysis of vessel movements and strategies using commercial catch and effort data from the New Zealand hoki fishery”. In: *Canadian Journal of Fisheries and Aquatic Sciences* 53.9 (1996), pp. 2126–2136.

[267] Tjeerd de Vries and Alexis Akira Toda. “Capital and Labor Income Pareto Exponents across Time and Space”. In: *arXiv:2006.03441 [econ, q-fin]* (2020). arXiv: 2006.03441. URL: http://arxiv.org/abs/2006.03441.

[268] H. W. Watson and Francis Galton. “On the Probability of the Extinction of Families.” In: *The Journal of the Anthropological Institute of Great Britain and Ireland* 4 (1875), p. 138. ISSN: 09595295. DOI: 10.2307/2841222.

[269] Duncan J. Watts and Steven H. Strogatz. “Collective dynamics of ‘small-world’ networks”. In: *Nature* 393.6684 (June 1998), pp. 440–442. DOI: 10.1038/30918. URL: https://doi.org/10.1038/30918.

[270] F. Wegner. “Inverse participation ratio in $2 + \varepsilon$ dimensions”. In: *Zeitschrift für Physik B Condensed Matter and Quanta* 36.3 (Sept. 1980), pp. 209–214. DOI: 10.1007/bf01325284. URL: https://doi.org/10.1007/bf01325284.

[271] G. Weisbuch, G. Deffuant, F. Amblard, and J.-P. Nadal. “Interacting Agents and Continuous Opinions Dynamics”. In: *Heterogenous Agents, Interactions and Economic Performance*. Ed. by Robin Cowan and Nicolas Jonard. Lecture Notes in Economics and Mathematical Systems. Springer, 2003, pp. 225–242. ISBN: 978-3-642-55651-7. DOI: 10.1007/978-3-642-55651-7_14.

[272] Eugene P. Wigner. “Random Matrices in Physics”. In: *SIAM Review* 9.1 (1967), pp. 1–23. DOI: 10.1137/1009001. URL: https://doi.org/10.1137%2F1009001.

[273] Eugene Paul Wigner. *Statistical properties of real symmetric matrices with many dimensions*. Princeton University, 1957.

[274] Michael A. Williams, Brijesh P. Pinto, and David Park. “Global evidence on the distribution of firm growth rates”. In: *Physica A: Statistical Mechanics and its Applications* 432 (2015), pp. 102–107. ISSN: 0378-4371. DOI: 10.1016/j.physa.2015.02.103.

[275] J. C. Willis. *Age and area; a study in geographical distribution and origin of species,* The University Press, 1922. URL: https://www.biodiversitylibrary.org/bibliography/30741.

[276] J. C. Willis and G. Udny Yule. “Some Statistics of Evolution and Geographical Distribution in Plants and Animals, and their Significance”. In: *Nature* 109.2728 (1922), pp. 177–179. ISSN: 1476-4687. DOI: 10.1038/109177a0.

[277] Charlie Wood. *The Cartoon Picture of Magnets That Has Transformed Science*. URL: https://www.quantamagazine.org/the-cartoon-picture-of-magnets-that-has-transformed-science-20200624/.

[278] Sewall Wright. “Statistical genetics and evolution”. In: *Bull. Amer. Math. Soc.* 48.4 (Apr. 1942), pp. 223–246. URL: https://projecteuclid.org:443/euclid.bams/1183504254.

[279] Matthieu Wyart and Jean-Philippe Bouchaud. “Statistical models for company growth”. In: *Physica A: Statistical Mechanics and its Applications* 326.1-2 (2003), pp. 241–255. DOI: 10.1016/s0378-4371(03)00267-x. URL: https://doi.org/10.1016%2Fs0378-4371%2803%2900267-x.

[280] George Udny Yule. "II.—A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S". In: *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* 213.402-410 (1925), pp. 21–87. DOI: 10.1098/rstb.1925.0002.

[281] Stefano Zapperi, Kent Bækgaard Lauritsen, and H. Eugene Stanley. "Self-Organized Branching Processes: Mean-Field Theory for Avalanches". In: *Physical Review Letters* 75.22 (1995), pp. 4071–4074. DOI: 10.1103/PhysRevLett.75.4071.

[282] Gabriel Zucman. "Global Wealth Inequality". In: (2019), p. 33.